Tight Bounds for Randomized Load Balancing on Arbitrary Network Topologies

Thomas Sauerwald He Sun
Max Planck Institute for Informatics
Saarbrücken, Germany
{sauerwal, hsun}@mpi-inf.mpg.de

Abstract

We consider the problem of balancing load items (tokens) on networks. Starting with an arbitrary load distribution, we allow in each round nodes to exchange tokens with their neighbors. The goal is to achieve a distribution where all nodes have nearly the same number of tokens.

For the continuous case where tokens are arbitrarily divisible, most load balancing schemes correspond to Markov chains whose convergence is fairly well-understood in terms of their spectral gap. However, in many applications load items cannot be divided arbitrarily and we need to deal with the discrete case where the load is composed of indivisible tokens. This discretization entails a non-linear behavior due to its rounding errors, which makes the analysis much harder than in the continuous case. Therefore, it has been a major open problem to understand the limitations of discrete load balancing and its relation to the continuous case.

We investigate several randomized protocols for different communication models in the discrete case. Our results demonstrate that there is almost no difference between the discrete and continuous case. For instance, for any regular network in the matching model, all nodes have the same load up to an additive constant in (asymptotically) the same number of rounds required in the continuous case. This generalizes and tightens the previous best result, which only holds for expander graphs (STOC'09).

Keywords: randomized algorithms, parallel and distributed algorithms, graph expansion, Markov chains, load balancing.

Contents

1 Introduction

Consider an application running on a parallel network with n processors. Every processor has initially a certain amount of *tokens* (tasks) and the processors are connected by an arbitrary graph. The goal of load balancing is to reallocate the tokens by transferring them along the edges so that eventually every processor has almost the same number of tokens.

Load balancing is a well-studied problem in distributed systems and has manifold applications in scheduling [39], hashing [28], routing [12], numerical computation such as solving partial differential equations [38, 40, 42] and simulating dynamics [10]. This trend has been reinforced by the flattening of processor speeds leading to an increasing usage of multi-core processors [6, 25] and the emergence of large decentralized networks like P2P networks [2, 21, 39]. Especially for large-scale networks, it is desirable to use local and iterative load balancing protocols, where every processor only needs to know its current and the neighboring processors' loads and based on this decides how many tokens should be sent (or received).

A widely used approach is the so-called diffusion (i.e., the first-order-diffusion scheme [12, 32]), where the amount of load sent along each edge in each round is proportional to the load difference between the incident nodes. The alternative is the matching model where in each round there is a matching and only those edges can be used for averaging the load.

We measure the smoothness of the load distribution by the so-called *discrepancy* which is the difference between the maximum and minimum load among all nodes. In view of more complex scenarios where jobs are eventually removed or new jobs are generated, the discrepancy seems to be a more appropriate measure than the *makespan*, which only considers the maximum load.

Many studies on load balancing assume that the load is arbitrarily divisible. In this *continuous case*, the diffusion scheme corresponds to a Markov chain on the graph and one can resort to a battery of established techniques to analyze the convergence speed [9, 19, 32]. In particular, the *spectral gap* captures the time to reach a small discrepancy quite accurately [35, 37]. This relation continues to hold for the matching model, even if the matchings are generated randomly, which might be necessary for graphs with no canonical matchings [11, 31].

However, in many applications a processors' load may consist of tasks which are not further divisible, which is why the continuous case is also referred to as "idealized case" [35]. A common way to model indivisible tasks is the *unit-size token model* where one assumes a smallest load entity, the unit-size token, and load is always represented by a multiple of this smallest entity. In the following, we will refer to the unit-size token model as the *discrete case*. Because of the close relation between continuous load balancing and Markov chains, many authors [19, 26, 31, 32, 35, 38] asked for a characterization of the convergence speed of discrete load balancing, or alternatively, a quantification of the deviation between the discrete and the continuous case. Unfortunately, the discrete case is much harder to analyze due to its nonlinearity caused by the roundings to whole tokens in each round.

Muthukrishnan et al. [32] proved the first rigorous result for the discrete case in the diffusion model. They assume that the load amount sent along each edge is obtained by rounding down the load amount that would be sent in the continuous case. Using this approach, they showed that the discrepancy is at most $\mathcal{O}\big(\frac{dn}{1-\lambda}\big)$ after $\mathcal{O}\big(\frac{\log(Kn)}{1-\lambda}\big)$ rounds, where d is the degree, K is the discrepancy of the initial load vector and $1-\lambda$ is the spectral gap of the diffusion matrix. Similar results for the matching model were shown by Muthukrishnan and Ghosh [31].

Further progress was made by Rabani et al. [35] who introduced the so-called *local divergence*, which is a natural parameter that essentially aggregates the sum of load differences over all edges in all rounds. For both the diffusion and matching model, they proved that the local divergence yields an upper bound on the maximum deviation between the continuous and discrete case for the aforementioned rounding down approach. They also computed the local divergence for different networks such as torus graphs and proved a general upper bound which translates into a discrepancy bound of $\mathcal{O}\big(\frac{d\log n}{1-\lambda}\big)$ after $\mathcal{O}\big(\frac{\log(Kn)}{1-\lambda}\big)$ rounds for any d-regular graph.

While always rounding down may lead to a quick stabilization, the discrepancy could be quite large, i.e., as large as the diameter of the graph (in case of diffusion, it could be even the diameter times the degree). Therefore, Rabani et al. [35] also suggested to use randomized rounding in order to get a better approximation of the continuous case. Herlihy and Tirthapura [20] analyzed such a protocol for the hypercube in the matching model and proved a discrepancy bound of $\mathcal{O}(\sqrt{\log n})$ after $\log_2 n$ rounds. Friedrich and Sauerwald [17] presented the first general analysis of this randomized protocol in the matching model. By analyzing the ℓ_2-version of the local divergence, the so-called *local* 2*-divergence*, they proved that on many networks, the randomized protocol yields a square root improvement in terms of the achieved discrepancy compared to the deterministic protocol from [35].

Recently, Berenbrink et al. [8] extended some of the results from [17] to the diffusion model. One general challenge in the diffusion model is that nodes may receive too many (or too few) tokens in a single round, since all neighbors have to make their decisions locally and independent of each other. This might explain why most discrepancy bounds for diffusion depend on the degree of the network and are thus weaker than the corresponding bounds for the matching model. Additionally, also practical simulations seem to favor the matching models, especially for fine load balancing [41].

Closely related to our problem addressed here are balancing networks [5], which are siblings of sorting networks with comparators replaced by balancers. Klugerman and Plaxton [23] gave the first construction of a balancing network of depth $\mathcal{O}(\log n)$ which achieves a discrepancy of one. Their network relies on the famous AKS sorting network [4]. Rabani et al. [35] derived results for other networks, but these involve a much larger depth. All of these results [5, 23, 35] require each balancer to be initialized in a special way, while our randomized protocols do not require any specific initialization and therefore seem to be more practical.

There are also studies in which the nodes are equipped with additional abilities compared to our model. For instance, Even-Dar and Mansour [16] analyzed a load balancing model where every node knows the average load. Elsässer and Sauerwald [15] analyzed an algorithm which uses random-walk based routing of positive and negative tokens to minimize the makespan.

While all aforementioned load balancing protocols can send an arbitrary number of tokens along the used edges, several studies consider an alternative model in which only a single token can traverse along each edge in each round. Obviously, the convergence is much slower and at least linear in the initial discrepancy (for concrete results, see Aiello et al. [3], Ghosh et al. [19]).

Finally, many (re-)allocation schemes have been analyzed which are based on the famous power-of-two choices paradigm for balls-and-bins models (cf. Mitzenmacher [30]). While there are results in distributed settings (e.g., Adler et al. [1], Lenzen and Wattenhofer [24]), most of them assume a complete graph as the underlying network. One notable exception is the work of Kenthapadi and Panigrahy [22]. However, they only considered the sequential allocation of n tokens and required the degree to be at least polynomial in n in order to achieve a discrepancy of $\mathcal{O}(\log\log n)$.

Our Results. We analyze several natural randomized protocols for indivisible tokens. All protocols have in common that randomized rounding is used to "imitate" the behavior of the continuous case in each round. Our main result for the matching model is as follows:

Theorem 1.1. *Let G be a regular graph with n nodes and K be the discrepancy of the initial load vector. There is a constant $c > 0$ independent of G and K, so that w.p. $1 - \exp(-(\log n)^{\Omega(1)})$, the discrepancy is at most c after $\mathcal{O}\left(\frac{\log(Kn)}{1-\lambda(\mathbf{P})}\right)$ rounds in the random matching model. This also holds after $\mathcal{O}\left(\frac{\log(Kn)}{1-\lambda(\mathbf{M})}\right)$ rounds in the balancing circuit model if d is constant.* [1]

[1]For precise definitions of both models, $\lambda(\mathbf{P})$, $\lambda(\mathbf{M})$ and d, we refer to Section 2.

Graph Family	Rounds	Discrepancy	Model	Ref.
Constant-Degree Expander Graphs	$\mathcal{O}(\log(Kn))$	$\mathcal{O}(\log n)$	det. (BC)	[35]
		$\mathcal{O}(\log\log n)$	rand. (BC & RM)	[17]
		$\mathcal{O}(1)$	rand. (BC & RM)	Thm. 1.1
r-dim. Torus Graphs	$\mathcal{O}\big(\log(Kn)\, n^{2/r}\big)$	$\mathcal{O}(n^{1/r})$	det. (BC)	[35]
		$\mathcal{O}(n^{1/(2r)}\sqrt{\log n})$	rand. (BC)	[17]
		$\mathcal{O}(n^{1/(2r)}\log n)$	rand. (RM)	[17]
		$\mathcal{O}(1)$	rand. (BC & RM)	Thm. 1.1
Regular Graphs	$\mathcal{O}\left(\frac{\log(Kn)}{1-\lambda}\right)$	$\mathcal{O}\left(\frac{d\log n}{1-\lambda}\right)$	det. (BC)	[35]
		$\mathcal{O}\left(\sqrt{\frac{d\log n}{1-\lambda}}\right)$	rand. (BC)	[17]
		$\mathcal{O}(1)$	rand. (BC, $d=\mathcal{O}(1)$)	Thm. 1.1
		$\mathcal{O}\left(\sqrt{\frac{(\log n)^3}{1-\lambda}}\right)$	rand. (RM)	[17]
		$\mathcal{O}(1)$	rand. (RM)	Thm. 1.1
Arbitrary Graphs	$\mathcal{O}\left(\frac{d\cdot\log(Kn)}{1-\lambda}\right)$	$\mathcal{O}\left(\frac{d\cdot\log n}{1-\lambda}\right)$	det. (BC)	[35]
		$\mathcal{O}\left(\sqrt{\frac{d\cdot\log n}{1-\lambda}}\right)$	rand. (BC)	[17]
	$\mathcal{O}(\tau_{\text{cont}}(K,n^{-2}))$	$\mathcal{O}((\log n)^{\varepsilon})$	rand. (BC & RM)	Thm. 3.7
	$\mathcal{O}(\tau_{\text{cont}}(K,n^{-2})\cdot(\log\log n))$	$\mathcal{O}(\log\log n)$	rand. (BC & RM)	Thm. 3.7

Table 1: Comparison of the results for the matching model with the previously best results. The initial discrepancy is denoted by K, and $1-\lambda$ denotes the spectral gap. Here, det. and rand. refer to the deterministic and randomized orientation, respectively. BC (RM) stand for the balancing circuit (random matching) model, respectively. Note that $\tau_{\text{cont}}(K,n^{-2})$ is the time for the continuous process to reach a discrepancy of n^{-2} w.p. $1-n^{-1}$. For the precise definitions, see Section 2.

The two bounds on the runtime in Theorem 1.1 match the ones from the continuous case up to a constant factor (see Theorem 2.2 and Theorem 2.5). The previous best result for this protocol holds *only* for expander graphs and the number of rounds is a factor $(\log\log n)^3$ larger than ours [17]. For expander graphs and $K=\text{poly}(n)$, our algorithm needs only $\Theta(\log n)$ rounds, which would be even necessary for any centralized algorithm. For general graphs, all previous bounds on the discrepancy include the spectral gap $1-\lambda$. Therefore, especially for graphs which have small expansion like Torus graphs, our main result represents a vast improvement (Table 1).

We further analyze the matching model on non-regular graphs and our result (Theorem 3.7) is almost tight, since the discrepancy is $\mathcal{O}(\log\log n)$ and the runtime is only an $\mathcal{O}(\log\log n)$ factor larger than in the continuous case. Together with Theorem 1.1, these results show that for *arbitrary* networks, there is almost no difference between the discrete and continuous case.

Finally, we also study two natural diffusion-based protocols in the discrete case [8, 18]. Our discrepancy bounds there depend only polynomially on the maximum degree Δ and logarithmically on n, while again all previous results include the spectral gap or are restricted to special graph classes [8, 18, 32, 35].

Our Techniques. Our main results are based on the combination of two novel techniques which may have further applications to other problems. First, instead of analyzing the rounding errors for each edge directly [8, 17, 31, 32, 35], we adopt a token-based viewpoint and relate the movements of tokens to independent random walks. This establishes a nice analogy between the distribution of tokens and the well-studied balls-and-bins model (see Corollary 3.4). Secondly,

we employ potential functions to reduce the task of balancing an arbitrary load vector to the task of balancing a *sparse* load vector, i.e., a load vector that contains much fewer tokens than n. Especially for these sparse load vectors, the token-based viewpoint yields much stronger concentration inequalities than the ones from previous work.

All of our discrepancy bounds make use of the so-called local 2-divergence, which has been one of the most important tools to quantify the deviation between the continuous and the discrete case [8, 17, 35]. We prove that for *any* graph and *any* sequence of matchings, the local 2-divergence is between 1 and $\sqrt{2}$, while all previous bounds on the local divergence include graph parameters such as the spectral gap or the (maximum) degree. For the diffusion model, the local 2-divergence is essentially given by $\Theta(\sqrt{\Delta})$, where Δ is the maximum degree. Prior to this, all bounds on the local 2-divergence in both communication models depend on the size and expansion of the graph, or are restricted to certain graph classes.

Organization. The remainder of this paper is organized as follows. Section 2 introduces the matching-based model and presents some basic results. In Section 3 we introduce our new technique that relates the movement of the tokens to independent random walks. Based on this technique, we derive results on the discrepancy that hold for arbitrary graphs (see Section 3.2). The proof of our main result (Theorem 1.1) is given in Section 4. Finally, Section 5 contains our results for the diffusion model.

Notations. We assume that $G = (V, E)$ is an undirected, connected graph with n nodes, indexed from 1 to n. Several inequalities in this paper require that n is sufficiently large. For any node u, let $N(u)$ be the set of neighbors of node u and $d(u) := |N(u)|$ the degree of node u. The maximum degree of G is denoted by $\Delta := \max_u d(u)$. By $\mathrm{diam}(G)$ we denote the diameter of G. Following [35], we use the notation $[u : v]$ for an edge $\{u, v\} \in E$ with $u < v$. For any vector $\mathbf{x} = (x_1, \ldots, x_n)$, the p-norm of $\mathbf{x}$ is defined by $\|\mathbf{x}\|_p := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$. In particular, $\|\mathbf{x}\|_\infty := \max_{1 \leqslant i \leqslant n} |x_i|$. For any n by n real symmetric matrix $\mathbf{M}$, let $\lambda_1(\mathbf{M}) \geqslant \ldots \geqslant \lambda_n(\mathbf{M})$ be the n eigenvalues of matrix $\mathbf{M}$. For simplicity, let $\lambda(\mathbf{M}) := \max\{|\lambda_2(\mathbf{M})|, |\lambda_n(\mathbf{M})|\}$. By $\log(\cdot)$ we denote the natural logarithm.

2 The Matching Model

In the *matching model* (also known as *dimension exchange model* reflecting its seminal application to hypercubes), every two matched nodes in round t balance their loads as evenly as possible. This can be expressed by a symmetric n by n matching matrix $\mathbf{M}^{(t)}$, where with slight abuse of notation we use the same symbol for the matching and the corresponding matching matrix. Matrix $\mathbf{M}^{(t)}$ is defined by $\mathbf{M}^{(t)}_{u,u} := 1/2$, $\mathbf{M}^{(t)}_{v,v} := 1/2$ and $\mathbf{M}^{(t)}_{u,v} = \mathbf{M}^{(t)}_{v,u} := 1/2$ if $\{u, v\} \in \mathbf{M}^{(t)} \subseteq E$ and $\mathbf{M}^{(t)}_{u,u} = 1$, $\mathbf{M}^{(t)}_{u,v} = 0$ $(u \neq v)$ if u is not matched. We will often consider the product of consecutive matching matrices and denote this by $\mathbf{M}^{[t_1,t_2]} = \prod_{s=t_1}^{t_2} \mathbf{M}^{(s)}$ for two rounds $t_1 \leqslant t_2$. If $t_1 \geqslant t_2 + 1$, then $\mathbf{M}^{[t_1,t_2]}$ is defined as the n by n identity matrix $\mathbf{I}$.

2.1 Balancing Circuit and Random Matching Model

In the *balancing circuit model*, a certain sequence of matchings is applied periodically. More precisely, let $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(d)}$ be a sequence of d matching matrices[2]. Then in round $t \geqslant 1$, we apply the matching matrix $\mathbf{M}^{(t)} := \mathbf{M}^{(((t-1) \bmod d)+1)}$. Following [35], we define the *round matrix* $\mathbf{M} := \prod_{s=1}^{d} \mathbf{M}^{(s)}$. Further, let $\lambda(\mathbf{M}) := \max\{|\lambda_2(\mathbf{M})|, |\lambda_n(\mathbf{M})|\}$. We always assume

[2] Traditionally, the variable d has been used for the number of matchings [17, 35]. There may not exist a direct relation between d and the (maximum) degree of the underlying graph G. However, the graph *induced* by the union of the d matchings has maximum degree at most d.

that $\lambda(\mathbf{M}) < 1$ which is equivalent to the matrix $\mathbf{M}$ being ergodic. A natural choice for the d matching matrices is given by an edge coloring of graph G. There are various efficient distributed edge coloring algorithms (see for example, Panconesi and Srinivasan [33, 34]).

The alternative to the balancing circuit model is the *random matching model*, where one generates a random matching in each round. There are several simple and distributed randomized protocols to generate such matchings. For instance, [31] analyzed a two-stage protocol for d-regular graphs where in the first stage every edge is picked independently with probability $\Theta(1/d)$. In the second stage, we consider the matching formed by all edges that are not incident to any other edge chosen in the first stage. A similar protocol was studied in [11] which also works for non-regular graphs. These protocols have two natural properties which are sufficient for our analysis. First, we have $p_{\min} = \Omega(1/\Delta)$, where $p_{\min} := \min_{t\in\mathbb{N}} \min_{\{u,v\}\in E} \mathbf{Pr}\left[\, \{u,v\} \in \mathbf{M}^{(t)} \,\right]$. Secondly, random matchings generated in different rounds are mutually independent.

2.2 The Continuous Case

In the continuous case the load is arbitrarily divisible. Let $\xi^{(0)}$ be the initial load vector and in every round two matched nodes balance their loads perfectly. It is easy to see that this process corresponds to a linear system and the load vector $\xi^{(t)}$, $t \geqslant 1$, can be expressed as $\xi^{(t)} = \xi^{(t-1)}\,\mathbf{M}^{(t)}$, which results in $\xi^{(t)} = \xi^{(0)}\,\mathbf{M}^{[1,t]}$. Moreover,

$$\begin{aligned}
\xi_u^{(t)} &= \xi_u^{(t-1)} + \sum_{v\colon \{u,v\}\in E} \left(\xi_v^{(t-1)} \mathbf{M}_{v,u}^{(t)} - \xi_u^{(t-1)} \mathbf{M}_{u,v}^{(t)} \right) \\
&= \xi_u^{(t-1)} + \sum_{v\colon \{u,v\}\in \mathbf{M}^{(t)}} \left(\frac{1}{2}\xi_v^{(t-1)} - \frac{1}{2}\xi_u^{(t-1)} \right) .
\end{aligned}$$

We define the average load by $\overline{\xi} := \sum_{w\in V} \xi_w^{(0)}/n$, which is invariant of the round t. Note that the convergence in the continuous case depends on the randomly chosen matchings in the random matching model, while it is "deterministic" in the balancing circuit model (for fixed initial load vector).

Definition 2.1. *Let G be any graph. Fix any pair (K,ε) with $K \geqslant \varepsilon > 0$. For any pair of integers $t_1 < t_2$, we call a time-interval $[t_1,t_2]$ associated with a sequence of matchings $\langle \mathbf{M}^{(t_1+1)}, \ldots, \mathbf{M}^{(t_2)} \rangle$ (K,ε)–smoothing if for any $\xi^{(t_1)} \in \mathbb{R}^n$, $\operatorname{disc}\left(\xi^{(t_1)}\right) \leqslant K \Rightarrow \operatorname{disc}\left(\xi^{(t_2)}\right) \leqslant \varepsilon$, where the discrepancy of any load vector ξ is defined by $\operatorname{disc}(\xi) := \max_{u,v\in V} |\xi_u - \xi_v|$.*

- *For the balancing circuit model, define*

$$\tau_{\text{cont}}(K,\varepsilon) := \min\left\{ t \in \mathbb{N}\colon [0,t] \text{ is } (K,\varepsilon)\text{–smoothing} \right\},$$

i.e., $\tau_{\text{cont}}(K,\varepsilon)$ is the minimum number of rounds in the continuous case to reach discrepancy ε for any initial vector $\xi^{(0)}$ with discrepancy at most K.

- *For the random matching model, define*

$$\tau_{\text{cont}}(K,\varepsilon) := \min\left\{ t \in \mathbb{N}\colon \mathbf{Pr}\left[\, [0,t] \text{ is } (K,\varepsilon)\text{–smoothing} \,\right] \geqslant 1 - n^{-1} \right\},$$

i.e., $\tau_{\text{cont}}(K,\varepsilon)$ is the minimum number of rounds in the continuous case so that with probability at least $1-n^{-1}$, we reach a discrepancy of ε for any initial vector $\xi^{(0)}$ with discrepancy at most K. Note that the probability space is taken over the t randomly chosen matchings $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(t)}$.

Note that in both the balancing circuit and the random matching model, $\tau_{\text{cont}}(K,\varepsilon)$ is *not* a random variable.

Following previous works [32, 35], we adopt the view that the continuous case $(\tau_{\text{cont}}(K,\varepsilon))$ is well-understood and our goal is to analyze the discrete case. For the balancing circuit model, there is indeed a natural bound on $\tau_{\text{cont}}(K,\varepsilon)$ depending on the spectral gap of $\mathbf{M}$.

Theorem 2.2 ([35, Theorem 1])**.** *Let G be any graph. Consider the balancing circuit model with d matchings $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(d)}$. Then for any $\varepsilon > 0$, $\tau_{\mathrm{cont}}(K, \varepsilon) \leqslant d \cdot \frac{4}{1-\lambda(\mathbf{M})} \cdot \log\left(\frac{Kn}{\varepsilon}\right)$.*

For the random matching model, our result will depend on $p_{\min}$ and the spectral gap of the *diffusion matrix* $\mathbf{P}$, defined as $\mathbf{P}_{u,v} := \frac{1}{2\Delta}$ if $\{u, v\} \in E$, $\mathbf{P}_{u,v} := 1 - \frac{d(u)}{2\Delta}$ if $v = u$, and $\mathbf{P}_{u,v} := 0$ otherwise.

Theorem 2.3 ([31, Theorem 1])**.** *Let G be any d-regular graph. Consider the random matching model in the continuous case and let $p_{\min}$ be the minimum probability for an edge to be included in the matching. Define the quadratic potential as $\Phi^{(t)} := \sum_{u \in V} \left(\xi_u^{(t)} - \overline{\xi}\right)^2$. Then for any round t, we have*

$$\mathbf{E}\left[\Phi^{(t)}\right] \leqslant \left(1 - \frac{d \cdot p_{\min}}{4} \cdot (1 - \lambda_2(\mathbf{P}))\right)^t \cdot \Phi^{(0)}.$$

Theorem 2.3 implies the following corollary.

Corollary 2.4. *For any node $v \in V$ and round t, it holds for a constant $c > 0$ that*

$$\mathbf{Pr}\left[\left\|\mathbf{M}_{v,\cdot}^{[1,t]}\right\|_2 \leqslant \frac{1}{n} + \mathrm{e}^{-c \cdot (1-\lambda_2) \cdot t}\right] \geqslant 1 - \mathrm{e}^{-c \cdot (1-\lambda_2) \cdot t}.$$

Proof. Let $\xi^{(0)}$ be the initial load vector with $\xi_v^{(0)} := 1$ and $\xi_u^{(0)} := 0$ for $u \neq v$. Then,

$$\xi_u^{(t)} = \sum_{w \in V} \xi_w^{(0)} \cdot \mathbf{M}_{w,u}^{[1,t]} = 1 \cdot \mathbf{M}_{v,u}^{[1,t]},$$

and $\overline{\xi} = 1/n$. By Theorem 2.3,

$$\mathbf{E}\left[\Phi^{(t)}\right] \leqslant \left(1 - \frac{d}{4} \cdot (1 - \lambda_2(\mathbf{P})) \cdot p_{\min}\right)^t \cdot \Phi^{(0)}.$$

Since $p_{\min} = \Omega(1/d)$, let $p_{\min} \geqslant c'/d$ for a constant $c' \in \mathbb{R}$. Using $\Phi^{(0)} = \left(1 - \frac{1}{n}\right)^2 + (n-1) \cdot \left(\frac{1}{n}\right)^2 = 1 - \frac{1}{n} \leqslant 1$, we obtain

$$\mathbf{E}\left[\Phi^{(t)}\right] \leqslant \left(1 - \frac{c'}{4} \cdot (1 - \lambda_2(\mathbf{P}))\right)^t \cdot 1 \leqslant \mathrm{e}^{-c'/4 \cdot (1-\lambda_2) \cdot t}.$$

Hence by Markov's inequality,

$$\mathbf{Pr}\left[\Phi^{(t)} \geqslant \mathrm{e}^{-c'/8 \cdot (1-\lambda_2) \cdot t}\right] \leqslant \mathrm{e}^{-c'/8 \cdot (1-\lambda_2) \cdot t}.$$

Assuming that $\Phi^{(t)} \leqslant \mathrm{e}^{-c'/8 \cdot (1-\lambda_2) \cdot t}$ implies for any node $u \in V$,

$$\left(\mathbf{M}_{v,u}^{[1,t]} - \frac{1}{n}\right)^2 = \left(\xi_u^{(t)} - \frac{1}{n}\right)^2 \leqslant \mathrm{e}^{-c'/8 \cdot (1-\lambda_2) \cdot t},$$

and rearranging yields

$$\mathbf{M}_{v,u}^{[1,t]} \leqslant \frac{1}{n} + \mathrm{e}^{-c'/16 \cdot (1-\lambda_2) \cdot t}.$$

By setting $c := c'/16$ we finish the proof. ∎

Theorem 2.3 also implies the following upper bound on $\tau_{\mathrm{cont}}(K, \varepsilon)$.

Theorem 2.5. *Let G be any d-regular graph and consider the random matching model. Then for any $\varepsilon > 0$, it holds that*

$$\tau_{\mathrm{cont}}(K, \varepsilon) \leqslant \frac{8}{d \cdot p_{\min}} \cdot \frac{1}{1 - \lambda_2(\mathbf{P})} \cdot \log\left(\frac{Kn^2}{\varepsilon/2}\right).$$

Hence for $p_{\min} = \Theta(1/d)$, we obtain essentially the same convergence as for the first-order-diffusion scheme (cmp. Theorem 5.1), although the communication is restricted to a single matching in each round. A more complicated but also more general result for non-regular graphs can be found in [11, Theorem 5]. However, since our proof for non-regular graphs does not require a concrete runtime bound, we prefer to state our results for non-regular graphs in terms of $\tau_{\mathrm{cont}}(K, \varepsilon)$.

Proof of Theorem 2.5. Fix any vertex $v \in V$ and let $\xi^{(0)}$ be the unit vector which has 1 at position v and 0 otherwise. Define the quadratic potential function as $\Phi^{(t)} = \sum_{u \in V} \left(\xi_u^{(t)} - \overline{\xi}\right)^2$. Hence, $\Phi^{(0)} = 1 - \frac{1}{n} \leqslant 1$. Choosing $t := \frac{4}{d \cdot p_{\min}} \cdot \frac{1}{1-\lambda_2(\mathbf{P})} \cdot \log\left(\frac{K^2 n^4}{\varepsilon^2/4}\right)$ in Theorem 2.3 yields

$$\mathbf{E}\left[\Phi^{(t)}\right] \leqslant \left(1 - \frac{d \cdot p_{\min}}{4} \cdot (1 - \lambda_2(\mathbf{P}))\right)^t \cdot \Phi^{(0)} \leqslant \mathrm{e}^{-\log(K^2 n^4) + \log(\varepsilon^2/4)} \cdot 1 = \frac{\varepsilon^2}{4K^2 n^4}.$$

Hence by Markov's inequality,

$$\mathbf{Pr}\left[\Phi^{(t)} \geqslant \frac{\varepsilon^2}{4K^2 n^2}\right] \leqslant \frac{1}{n^2}.$$

If $\Phi^{(t)} \leqslant \frac{\varepsilon^2}{4K^2n^2}$, then for every node $u \in V$, we have $\left|\xi_u^{(t)} - \overline{\xi}\right| \leqslant \sqrt{\frac{\varepsilon^2}{4K^2n^2}} = \frac{\varepsilon}{2Kn}$. Since $\overline{\xi} = \frac{1}{n}$ and $\xi_u^{(t)} = \mathbf{M}_{v,u}^{[1,t]}$, it follows that

$$\left|\mathbf{M}_{v,u}^{[1,t]} - \frac{1}{n}\right| \leqslant \frac{\varepsilon}{2Kn}.$$

Therefore,

$$\mathbf{Pr}\left[\forall u \in V\colon \left|\mathbf{M}_{v,u}^{[1,t]} - \frac{1}{n}\right| \leqslant \frac{\varepsilon}{2Kn}\right] \geqslant 1 - n^{-2}.$$

Replacing the initial load vector $\xi^{(0)}$ by the other unit vectors, repeating the above argument and taking the union bound gives

$$\mathbf{Pr}\left[\forall u \in V, \forall v \in V\colon \left|\mathbf{M}_{v,u}^{[1,t]} - \frac{1}{n}\right| \leqslant \frac{\varepsilon}{2Kn}\right] \geqslant 1 - n \cdot n^{-2} = 1 - n^{-1}.$$

For the remainder of the proof, assume that for all nodes $u, v \in V$, $\left|\mathbf{M}_{v,u}^{[1,t]} - \frac{1}{n}\right| \leqslant \frac{\varepsilon}{2Kn}$ holds. Then if we start with any initial load vector $\xi^{(0)}$ with discrepancy at most K, then

$$\begin{aligned}
\xi_u^{(t)} &= \sum_{v \in V} \xi_v^{(0)} \cdot \mathbf{M}_{v,u}^{(t)} \\
&= \sum_{v \in V} \overline{\xi} \cdot \mathbf{M}_{v,u}^{(t)} + \sum_{v \in V} (\xi_v^{(0)} - \overline{\xi}) \cdot \mathbf{M}_{v,u}^{(t)} \\
&= \overline{\xi} + \sum_{v \in V} (\xi_v^{(0)} - \overline{\xi}) \cdot \frac{1}{n} + \sum_{v \in V} (\xi_v^{(0)} - \overline{\xi}) \cdot \left(\mathbf{M}_{v,u}^{(t)} - \frac{1}{n}\right) \\
&\leqslant \overline{\xi} + \sum_{v \in V} K \cdot \frac{\varepsilon}{2Kn} \leqslant \overline{\xi} + \frac{\varepsilon}{2}.
\end{aligned}$$

Similarly, $\xi_u^{(t)} \geqslant \overline{\xi} - \varepsilon/2$. Hence the discrepancy of ξ at the end of round t is at most ε. ∎

2.3 The Discrete Case

Let us now turn to the discrete case with indivisible, unit-size tokens. Let $x^{(0)} \in \mathbb{Z}^n$ be the initial load vector with average load $\overline{x} := \sum_{w \in V} x_w^{(0)}/n$ and $x^{(t)}$ be the load vector at the end of round t. If the sum of tokens of two matched nodes is odd, we have to decide which of the matched nodes should get the excess token. To this end, we employ the so-called *random orientation* ([17, 35]) in the spirit of randomized rounding. More precisely, for any two matched nodes u and v in round t, node u gets $\left\lceil \frac{x_u^{(t-1)}+x_v^{(t-1)}}{2} \right\rceil$ or $\left\lfloor \frac{x_u^{(t-1)}+x_v^{(t-1)}}{2} \right\rfloor$ tokens, with probability 1/2 each. The remaining tokens are assigned to node v. We can also think of this as first assigning $\left\lfloor \frac{x_u^{(t-1)}+x_v^{(t-1)}}{2} \right\rfloor$ tokens to both u and v and then assigning the excess token (if there is one) to u or v with probability 1/2 each. We use a uniform random variable $\Phi_{u,v}^{(t)} \in \{-1, 1\}$ to specify the orientation for an edge $\{u, v\}$ in $\mathbf{M}^{(t)}$, i.e., indicating where the excess token (if any) is assigned to. If $\Phi_{u,v}^{(t)} = 1$, then the excess token is assigned to u and if $\Phi_{u,v}^{(t)} = -1$, then the excess token is assigned to v. Note that $\Phi_{u,v}^{(t)} = -\Phi_{v,u}^{(t)}$. Further, we point out that the deterministic orientation of Rabani et al. [35] corresponds to $\Phi_{u,v}^{(t)} = 1$ for $x_u^{(t-1)} \geqslant x_v^{(t-1)}$ and $\Phi_{u,v}^{(t)} = -1$ otherwise.

For every edge $\{u, v\}$ which is part of the matching $\mathbf{M}^{(t)}$, define the corresponding error term by

$$e_{u,v}^{(t)} := \frac{1}{2}\mathsf{Odd}(x_u^{(t-1)} + x_v^{(t-1)}) \cdot \Phi_{u,v}^{(t)},$$

where $\mathsf{Odd}(x) := x \bmod 2$. Moreover, for any round t we define an error vector $e^{(t)}$ with $e_u^{(t)} := \sum_{v\colon \{u,v\} \in \mathbf{M}^{(t)}} e_{u,v}^{(t)}$. With this notation, the load vector in round t is $x^{(t)} = x^{(t-1)}\mathbf{M}^{(t)} + e^{(t)}$. Solving this recursion (cf. [35]) yields

$$x^{(t)} = x^{(0)}\mathbf{M}^{[1,t]} + \sum_{s=1}^{t} e^{(s)}\mathbf{M}^{[s+1,t]} = \xi^{(t)} + \sum_{s=1}^{t} e^{(s)}\mathbf{M}^{[s+1,t]},$$

where $\xi^{(t)}$ is the corresponding load vector in the continuous case initialized with $\xi^{(0)} = x^{(0)}$. Hence, for any node $w \in V$ we have

$$x_w^{(t)} - \xi_w^{(t)} = \sum_{s=1}^{t} \sum_{u \in V} \sum_{v\colon \{u,v\} \in \mathbf{M}^{(s)}} e_{u,v}^{(s)} \mathbf{M}_{u,w}^{[s+1,t]} = \sum_{s=1}^{t} \sum_{[u:v] \in \mathbf{M}^{(s)}} e_{u,v}^{(s)} \left(\mathbf{M}_{u,w}^{[s+1,t]} - \mathbf{M}_{v,w}^{[s+1,t]} \right), \quad (2.1)$$

where the last equality used $e_{u,v}^{(s)} = -e_{v,u}^{(s)}$ (recall that $[u : v] \in \mathbf{M}^{(s)}$ means $\{u, v\} \in \mathbf{M}^{(s)}$ and $u < v$). Occasionally it will be convenient to "normalize" the load vector so that $\overline{x} \in [0, 1)$ (cmp. Observation 2.6). Although this may lead to negative entries in the load vector, the above formulas still hold.

Observation 2.6. *Fix a sequence of matchings $\mathcal{M} = \langle \mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \ldots \rangle$ and orientations $\Phi_{u,v}^{(t)}$, $[u : v] \in \mathbf{M}^{(t)}, t \in \mathbb{N}$. Consider two executions of the discrete load balancing protocol with the same matchings and orientations, but with different initial load vectors, $x^{(0)}$ and $\widetilde{x}^{(0)}$. Then the following statements hold:*

1. *If $\widetilde{x}^{(0)} = x^{(0)} + \alpha \cdot \mathbf{1}$ for some $\alpha \in \mathbb{Z}$, then $\widetilde{x}^{(t)} = x^{(t)} + \alpha \cdot \mathbf{1}$ for all $t \in \mathbb{N}$.*
2. *If $\widetilde{x}_u^{(0)} \leqslant x_u^{(0)}$ for all $u \in V$, then $\widetilde{x}_u^{(t)} \leqslant x_u^{(t)}$ for all $u \in V$ and $t \in \mathbb{N}$.*

The next lemma says that upper bounding the maximum load is essentially equivalent to lower bounding the minimum load.

Lemma 2.7. *Fix a sequence of matchings $\mathcal{M} = \langle \mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \ldots \rangle$. For any triple of non-negative integers K, α with $1 \leqslant \alpha \leqslant K$ and t,*

$$\max_{\substack{y \in \mathbb{Z}^n: \\ \mathrm{disc}(y) \leqslant K}} \left\{ \mathbf{Pr}\left[x_{\max}^{(t)} \geqslant \lfloor \overline{x} \rfloor + \alpha \,\middle|\, x^{(0)} = y \right] \right\} \leqslant \max_{\substack{y \in \mathbb{Z}^n: \\ \mathrm{disc}(y) \leqslant K}} \left\{ \mathbf{Pr}\left[x_{\min}^{(t)} \leqslant \lfloor \overline{x} \rfloor - \alpha + 3 \,\middle|\, x^{(0)} = y \right] \right\},$$

and similarly,

$$\max_{\substack{y \in \mathbb{Z}^n: \\ \mathrm{disc}(y) \leqslant K}} \left\{ \mathbf{Pr}\left[x_{\min}^{(t)} \leqslant \lfloor \overline{x} \rfloor - \alpha \,\middle|\, x^{(0)} = y \right] \right\} \leqslant \max_{\substack{y \in \mathbb{Z}^n: \\ \mathrm{disc}(y) \leqslant K}} \left\{ \mathbf{Pr}\left[x_{\max}^{(t)} \geqslant \lfloor \overline{x} \rfloor + \alpha - 3 \,\middle|\, x^{(0)} = y \right] \right\}.$$

Proof. We define a coupling between two executions of the load balancing algorithm. For every round t, the two executions use the same matching. In the first execution, we start with $x^{(0)} = z \in \mathbb{Z}^n$ that maximizes $\mathbf{Pr}\left[x_{\max}^{(t)} \geqslant \lfloor \overline{x} \rfloor + \alpha \,\middle|\, x^{(0)} = z \right]$ and satisfies $\mathrm{disc}(z) \leqslant K$. The load vector of the second execution is denoted by $\widetilde{x}^{(t)}$ and is initialized by

$$\widetilde{x}_u^{(0)} := \lfloor \overline{z} \rfloor - (z_u - \lfloor \overline{z} \rfloor) = 2 \cdot \lfloor \overline{z} \rfloor - z_u,$$

for any $u \in V$. Note that $\mathrm{disc}(\widetilde{x}^{(0)}) = \mathrm{disc}(z) \leqslant K$ and $\overline{\widetilde{x}} \geqslant \overline{x} - 2$. We couple the random choices of the two executions by setting for every $[u : v] \in \mathbf{M}^{(s)}$, $1 \leqslant s \leqslant t$, $\widetilde{\Phi}_{u,v}^{(s)} = -\Phi_{u,v}^{(s)}$, where $\widetilde{\Phi}$ denotes the random choices of the second execution. We now claim that for every round $s \geqslant 1$,

$$\widetilde{x}^{(s)} = 2 \cdot \lfloor \overline{x} \rfloor \cdot \mathbf{1} - x^{(s)}. \tag{2.2}$$

This claim is shown by induction on s as follows. First consider a node u which is matched with a node v in round s. Hence,

$$\widetilde{x}_u^{(s)} = \frac{1}{2} \cdot \widetilde{x}_u^{(s-1)} + \frac{1}{2} \cdot \widetilde{x}_v^{(s-1)} + \frac{1}{2} \cdot \mathsf{Odd}(\widetilde{x}_u^{(s-1)} + \widetilde{x}_v^{(s-1)}) \cdot \widetilde{\Phi}_{u,v}^{(s)},$$

and using the induction hypothesis yields

$$\begin{aligned}
\widetilde{x}_u^{(s)} &= \frac{1}{2} \cdot (2\lfloor x \rfloor - x_u^{(s-1)}) + \frac{1}{2} \cdot (2\lfloor x \rfloor - x_v^{(s-1)}) + \frac{1}{2} \cdot \mathsf{Odd}(\widetilde{x}_u^{(s-1)} + \widetilde{x}_v^{(s-1)}) \cdot \widetilde{\Phi}_{u,v}^{(s)} \\
&= 2\lfloor x \rfloor - \left(\frac{1}{2} \cdot x_u^{(s-1)} + \frac{1}{2} \cdot x_v^{(s-1)} + \frac{1}{2} \cdot \mathsf{Odd}(2\lfloor x \rfloor - x_u^{(s-1)} + 2\lfloor x \rfloor - x_v^{(s-1)}) \cdot \Phi_{u,v}^{(s)} \right) \\
&= 2\lfloor x \rfloor - x_u^{(s)},
\end{aligned}$$

where the last equality holds since $\mathsf{Odd}(2\lfloor x \rfloor - x_u^{(s-1)} + 2\lfloor x \rfloor - x_v^{(s-1)}) = \mathsf{Odd}(x_u^{(s-1)} + x_v^{(s-1)})$.

If a node u is not matched in round s, then the claim follows directly by the induction hypothesis. Hence, (2.2) holds which implies that for any $1 \leqslant \alpha \leqslant K$,

$$\begin{aligned}
\max_{\substack{y \in \mathbb{Z}^n: \\ \mathrm{disc}(y) \leqslant K}} \left\{ \mathbf{Pr}\left[x_{\max}^{(t)} \geqslant \lfloor \overline{x} \rfloor + \alpha \,\middle|\, x^{(0)} = y \right] \right\} &= \mathbf{Pr}\left[x_{\max}^{(t)} \geqslant \lfloor \overline{x} \rfloor + \alpha \,\middle|\, x^{(0)} = z \right] \\
&= \mathbf{Pr}\left[\widetilde{x}_{\min}^{(t)} \leqslant \lfloor \overline{x} \rfloor - \alpha \,\middle|\, \widetilde{x}^{(0)} = 2 \cdot \lfloor \overline{z} \rfloor \cdot \mathbf{1} - z \right] \\
&\leqslant \mathbf{Pr}\left[\widetilde{x}_{\min}^{(t)} \leqslant \lfloor \overline{(\widetilde{x})} \rfloor - \alpha + 3 \,\middle|\, \widetilde{x}^{(0)} = 2 \cdot \lfloor \overline{z} \rfloor \cdot \mathbf{1} - z \right] \\
&\leqslant \max_{\substack{y \in \mathbb{Z}^n: \\ \mathrm{disc}(y) \leqslant K}} \left\{ \mathbf{Pr}\left[x_{\min}^{(t)} \leqslant \lfloor \overline{x} \rfloor - \alpha + 3 \,\middle|\, x^{(0)} = y \right] \right\}.
\end{aligned}$$

The second inequality is shown in exactly the same way. This completes the proof. ■

Remark 2.8. *When referring to the random matching model (discrete case), the probability space is over the randomly generated matchings* and *the randomized orientation of the matchings. For the balancing circuit model (discrete case), the probability space is over the randomized orientation of the (deterministic) matchings.*

2.4 Local Divergence and Discrepancy

To bound the deviation between the discrete and continuous case, we consider $\max_{w \in V} \left| x_w^{(t)} - \xi_w^{(t)} \right|$ at all rounds t and define the local divergence for the matching model.

Definition 2.9 (Local p-Divergence for Matchings)**.** *For any graph G, $p \in \mathbb{Z}_+$ and an arbitrary sequence of matchings $\mathcal{M} = \langle \mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \dots \rangle$, the local p-divergence is*

$$\Psi_p(\mathcal{M}) = \max_{w \in V} \left(\sup_{t \in \mathbb{N}} \sum_{s=1}^{t} \sum_{[u:v] \in \mathbf{M}^{(s)}} \left| \mathbf{M}_{w,u}^{[s+1,t]} - \mathbf{M}_{w,v}^{[s+1,t]} \right|^p \right)^{1/p}.$$

Comparing the above definition with (2.1), one can see that $\Psi_1(\mathcal{M})$ *is a natural quantity that measures the sum of load differences across all edges in the network, aggregated over time [35]* and $\Psi_p(\mathcal{M})$ is the pth norm of $\Psi_1(\mathcal{M})$. Next we turn to our upper bound on the local 2-divergence.

Theorem 2.10. *For any graph G and any sequence of matchings $\mathcal{M}$, $\Psi_2(\mathcal{M}) \leqslant \sqrt{2 - 2 \cdot n^{-1}}$. Moreover, if there is a matching $\mathbf{M}^{(t)}$ in $\mathcal{M}$ such that the set $\mathbf{M}^{(t)} \neq \emptyset$, then $\Psi_2(\mathcal{M}) \geqslant 1$, otherwise $\Psi_2(\mathcal{M}) = 0$.*

While all previous upper bounds on the local divergence are functions of the expansion, the degree and/or the number of nodes [8, 17, 35], Theorem 2.10 establishes that the local 2-divergence is essentially independent of any graph parameter. Additionally, the local 1-divergence is always lower bounded by the diameter of graph G (cf. [17]).

Proof of Theorem 2.10. Fix any pair of node $w \in V$ and round t. For any $1 \leqslant s \leqslant t$, define the following potential function:

$$\Phi^{(s)} := \sum_{u \in V} \left(\mathbf{M}_{w,u}^{[s+1,t]} - \frac{1}{n} \right)^2.$$

Observe that since $\mathbf{M}^{[t+1,t]}$ is the identity matrix, $\Phi^{(t)} = 1 \cdot \left(1 - \frac{1}{n}\right)^2 + (n-1) \cdot \left(\frac{1}{n}\right)^2 = 1 - \frac{1}{n}$. Consider now any round $1 \leqslant s \leqslant t$, and let u, v be nodes with $[u:v] \in \mathbf{M}^{(s)}$. Let $y_u := \mathbf{M}_{w,u}^{[s+1,t]}$ and $y_v := \mathbf{M}_{w,v}^{[s+1,t]}$. Note that

$$\mathbf{M}_{w,u}^{[s,t]} = \sum_{k \in V} \mathbf{M}_{w,k}^{[s,s]} \cdot \mathbf{M}_{k,u}^{[s+1,t]} = \frac{y_u + y_v}{2},$$

and similarly, $\mathbf{M}_{w,v}^{[s,t]} = \frac{y_u + y_v}{2}$. Therefore, the contribution of u and v to $\Phi^{(s)} - \Phi^{(s-1)}$ is equal to

$$\begin{aligned}
&\left(\mathbf{M}_{w,u}^{[s+1,t]} - \frac{1}{n} \right)^2 + \left(\mathbf{M}_{w,v}^{[s+1,t]} - \frac{1}{n} \right)^2 - \left(\mathbf{M}_{w,u}^{[s,t]} - \frac{1}{n} \right)^2 - \left(\mathbf{M}_{w,v}^{[s,t]} - \frac{1}{n} \right)^2 \\
&= \left(y_u - \frac{1}{n} \right)^2 + \left(y_v - \frac{1}{n} \right)^2 - \left(\frac{y_u + y_v}{2} - \frac{1}{n} \right)^2 - \left(\frac{y_u + y_v}{2} - \frac{1}{n} \right)^2 \\
&= y_u^2 - \frac{2}{n} y_u + \frac{1}{n^2} + y_v^2 - \frac{2}{n} y_v + \frac{1}{n^2} - 2 \cdot \left(\frac{(y_u + y_v)^2}{4} - \frac{y_u + y_v}{n} + \frac{1}{n^2} \right) \\
&= y_u^2 + y_v^2 - \frac{y_u^2 + 2 y_u y_v + y_v^2}{2} \\
&= \frac{y_u^2}{2} - y_u y_v + \frac{y_v^2}{2} = \frac{1}{2} \cdot (y_u - y_v)^2.
\end{aligned}$$

If a node is not matched in round s, then its contribution to $\Phi^{(s)} - \Phi^{(s-1)}$ equals zero. Accumulating the contribution of all nodes yields

$$\Phi^{(s)} - \Phi^{(s-1)} = \sum_{[u:v] \in \mathbf{M}^{(s)}} \frac{1}{2} \cdot \left(\mathbf{M}^{[s+1,t]}_{w,u} - \mathbf{M}^{[s+1,t]}_{w,v} \right)^2 .$$

For the upper bound on $\Psi_2(\mathcal{M})$, we take the sum over t rounds to obtain that

$$\begin{aligned}\sum_{s=1}^{t} \sum_{[u:v] \in \mathbf{M}^{(s)}} \frac{1}{2} \cdot \left(\mathbf{M}^{[s+1,t]}_{w,u} - \mathbf{M}^{[s+1,t]}_{w,v} \right)^2 &= \sum_{s=1}^{t} \left(\Phi^{(s)} - \Phi^{(s-1)} \right) \\ &= \Phi^{(t)} - \Phi^{(0)} \leqslant 1 - \frac{1}{n},\end{aligned}$$

and thus

$$\sum_{s=1}^{t} \sum_{[u:v] \in \mathbf{M}^{(s)}} \left(\mathbf{M}^{[s+1,t]}_{w,u} - \mathbf{M}^{[s+1,t]}_{w,v} \right)^2 \leqslant 2 \cdot \left(1 - \frac{1}{n} \right),$$

which directly implies that $\Psi_2(\mathcal{M}) \leqslant \sqrt{2 - 2 \cdot n^{-1}}$. For the lower bound, consider any round t with $[u : v] \in \mathbf{M}^{(t)}$. Clearly,

$$\mathbf{M}^{[t+1,t]}_{u,u} = \mathbf{M}^{[t+1,t]}_{v,v} = 1, \quad \text{and} \quad \mathbf{M}^{[t+1,t]}_{u,v} = \mathbf{M}^{[t+1,t]}_{v,u} = 0,$$

and hence

$$\Psi_2(\mathcal{M}) \geqslant \sqrt{\sum_{[u:v] \in \mathbf{M}^{(t)}} \left(\mathbf{M}^{[s+1,t]}_{u,u} - \mathbf{M}^{[s+1,t]}_{u,v} \right)^2} \geqslant 1 . \qquad \blacksquare$$

Using the same arguments as in the proof of Theorem 2.10, we obtain the following corollary.

Corollary 2.11. *Let $G = (V, E)$ be any graph with an arbitrary sequence of matchings $\mathcal{M} = \langle \mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \ldots \rangle$. For an arbitrary node $w \in V$ and any pair of rounds $t_1 < t_2$, it holds that*

$$\sum_{s=1}^{t_1} \sum_{[u:v] \in \mathbf{M}^{(s)}} \left(\mathbf{M}^{[s+1,t_2]}_{w,u} - \mathbf{M}^{[s+1,t_2]}_{w,v} \right)^2 \leqslant 2 \cdot \sum_{u \in V} \left(\mathbf{M}^{[t_1+1,t_2]}_{w,u} - \frac{1}{n} \right)^2 .$$

Proof. The proof uses similar arguments as the proof of Theorem 2.10. Fix any node $w \in V$. For any $1 \leqslant s \leqslant t_2$, define $\Phi^{(s)} = \sum_{u \in V} \left(\mathbf{M}^{[s+1,t_2]}_{w,u} - \frac{1}{n} \right)^2$. As shown in the proof of Theorem 2.10,

$$\Phi^{(s)} - \Phi^{(s-1)} = \sum_{[u:v] \in \mathbf{M}^{(s)}} \frac{1}{2} \cdot \left(\mathbf{M}^{[s+1,t_2]}_{w,u} - \mathbf{M}^{[s+1,t_2]}_{w,v} \right)^2 .$$

Therefore,

$$\sum_{s=1}^{t_1} \sum_{[u:v] \in \mathbf{M}^{(s)}} \left(\mathbf{M}^{[s+1,t_2]}_{w,u} - \mathbf{M}^{[s+1,t_2]}_{w,v} \right)^2 = 2 \sum_{s=1}^{t_1} \left(\Phi^{(s)} - \Phi^{(s-1)} \right) = 2\Phi^{(t_1)} - 2\Phi^{(0)} \leqslant 2\Phi^{(t_1)},$$

which completes the proof. $\blacksquare$

Lemma 2.12. *Fix two rounds $t_1 < t_2$ and the load vector $x^{(t_1)}$ at the end of round t_1. For any family of non-negative numbers $g_{u,v}^{(s)}$ ($[u:v] \in \mathbf{M}^{(s)}, t_1+1 \leqslant s \leqslant t_2$), define the random variable Z by $Z := \sum_{s=t_1+1}^{t_2} \sum_{[u:v]\in\mathbf{M}^{(s)}} g_{u,v}^{(s)} \cdot e_{u,v}^{(s)}$. Then $\mathbf{E}[\,Z\,] = 0$ and for any $\delta > 0$ it holds that*

$$\mathbf{Pr}[\,|Z - \mathbf{E}[\,Z\,]| \geqslant \delta\,] \leqslant 2\exp\left(-\frac{\delta^2}{2\sum_{s=t_1+1}^{t_2}\sum_{[u:v]\in\mathbf{M}^{(s)}}\left(g_{u,v}^{(s)}\right)^2}\right).$$

Proof. The proof of this lemma is similar to [8, Proof of Theorem 1.1, first statement].

Since $\mathbf{E}\left[e_{u,v}^{(s)}\right] = 0$ for all $\{u,v\} \in \mathbf{M}^{(s)}$, it follows that $\mathbf{E}[\,Z\,] = 0$. Our goal is now to prove that Z is concentrated around its mean by applying the concentration inequality in Theorem A.1. Observe that Z depends on at most $(n/2)\cdot(t_2-t_1)$ random variables $e_{u,v}^{(s)}$, and each of them corresponds to one of the orientations of at most $n/2$ matching edges in each of the t_2-t_1 rounds. Let us denote this sequence by Y_ℓ with $(t_1+1)\cdot(n/2)+1 \leqslant \ell \leqslant (t_2+1)\cdot(n/2)$, where Y_ℓ with $\ell = \alpha\cdot(n/2)+\beta$, $t_1+1 \leqslant \alpha \leqslant t_2$, $1 \leqslant \beta \leqslant n/2$, describes the orientation of the β-th edge in round α (here we take an arbitrary ordering of the matching edges in round α, and if there are less than β matching edges in round α, then $Y_\ell \equiv 0$).

In order to apply Theorem A.1, we first verify that for every $(t_1+1)\cdot(n/2)+1 \leqslant \ell \leqslant (t_2+1)\cdot(n/2)$ with $\ell = \alpha\cdot(n/2)+\beta$ and $\{u',v'\}$ being the β-th matching edge in round α,

$$\left|\mathbf{E}\left[\,Z \mid Y_{(t_1+1)\cdot(n/2)+1},\dots,Y_\ell\,\right] - \mathbf{E}\left[\,Z \mid Y_{(t_1+1)\cdot(n/2)+1},\dots,Y_{\ell-1}\,\right]\right| \leqslant g_{u',v'}^{(\alpha)}. \tag{2.3}$$

In order to simplify the notation, let $\mathcal{Y}_\ell := (Y_{(t_1+1)\cdot(n/2)+1},\dots,Y_\ell)$ for any ℓ with $(t_1+1)\cdot(n/2)+1 \leqslant \ell \leqslant (t_2+1)\cdot(n/2)$.

To prove (2.3), we split the sum of Z into three parts: $s < \alpha$, $s = \alpha$ and $s > \alpha$.

Case 1: $t_1+1 \leqslant s \leqslant \alpha-1$**.** For every $[u:v] \in \mathbf{M}^{(s)}$, $e_{u,v}^{(s)}$ is determined by $\mathcal{Y}_{\ell-1}$. Hence,

$$\left|\mathbf{E}\left[\sum_{s=t_1+1}^{\alpha-1}\sum_{[u:v]\in\mathbf{M}^{(s)}} g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_\ell\right] - \mathbf{E}\left[\sum_{s=t_1+1}^{\alpha-1}\sum_{[u:v]\in\mathbf{M}^{(s)}} g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_{\ell-1}\right]\right| = 0.$$

Case 2: $s = \alpha$**.** Then,

$$\begin{aligned}
&\left|\mathbf{E}\left[\sum_{[u:v]\in\mathbf{M}^{(s)}} g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_\ell\right] - \mathbf{E}\left[\sum_{[u:v]\in\mathbf{M}^{(s)}} g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_{\ell-1}\right]\right| \\
&\leqslant \sum_{\substack{[u:v]\in\mathbf{M}^{(s)}\\ [u:v]\neq[u':v']}} \left|\mathbf{E}\left[g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_\ell\right] - \mathbf{E}\left[g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_{\ell-1}\right]\right| \\
&\qquad + \left|\mathbf{E}\left[g_{u',v'}^{(s)}\cdot e_{u',v'}^{(s)}\,\middle|\,\mathcal{Y}_\ell\right] - \mathbf{E}\left[g_{u',v'}^{(s)}\cdot e_{u',v'}^{(s)}\,\middle|\,\mathcal{Y}_{\ell-1}\right]\right| \\
&\leqslant g_{u',v'}^{(s)},
\end{aligned}$$

where the last inequality holds since $e_{u,v}^{(s)}$, $[u:v] \in \mathbf{M}^{(s)}, [u:v] \neq [u':v']$, are independent of $e_{u',v'}^{(s)}$ and $e_{u',v'}^{(s)} \in \{-1/2, 0, 1/2\}$.

Case 3: $\alpha+1 \leqslant s \leqslant t_2$**.** Let $\widetilde{\ell} \geqslant \ell$ be the smallest integer so that $\mathcal{Y}_{\widetilde{\ell}}$ determines the load vector $x^{(\alpha)}$. By the law of total expectation, for any $\{u,v\} \in \mathbf{M}^{(s)}$,

$$\mathbf{E}\left[\,e_{u,v}^{(s)} \mid \mathcal{Y}_\ell\,\right] = \mathbf{E}\left[\mathbf{E}\left[\,e_{u,v}^{(s)} \mid \mathcal{Y}_{\widetilde{\ell}}\,\right] \mid \mathcal{Y}_\ell\right] = \mathbf{E}[\,0 \mid \mathcal{Y}_\ell\,] = 0,$$

and the same also holds if we replace ℓ by $\ell-1$. Hence by linearity of expectation,

$$\mathbf{E}\left[\sum_{s=\alpha+1}^{t_2}\sum_{[u:v]\in\mathbf{M}^{(s)}} g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_\ell\right]-\mathbf{E}\left[\sum_{s=\alpha+1}^{t_2}\sum_{[u:v]\in\mathbf{M}^{(s)}} g_{u,v}^{(s)}\cdot e_{u,v}^{(s)}\,\middle|\,\mathcal{Y}_{\ell-1}\right]=0$$

Combining the contribution of all three cases establishes (2.3). Applying Theorem A.1 finishes the proof. ∎

We now list the following Chernoff-type bounds which can be derived quite easily from Lemma 2.12. Similar bounds have been derived in previous works [8, 17], but we obtain a much better concentration which is independent of the graph's expansion due to our new bound on the local 2-divergence (see the second statement of Lemma 2.13 below).

Lemma 2.13. *Fix an arbitrary load vector $x^{(0)}$. Consider two rounds $t_1 \leqslant t_2$ and assume that the time-interval $[0,t_1]$ is $(K,1/(2n))$–smoothing. Then for any node $k\in V$ and $\delta>1/n$, it holds that*

$$\mathbf{Pr}\left[\left|\sum_{w\in V} x_w^{(t_1)}\cdot\mathbf{M}_{w,k}^{[t_1+1,t_2]}-\overline{x}\right|\geqslant\delta\right]\leqslant 2\cdot\exp\left(-\frac{(\delta-1/(2n))^2}{4\sum_{w\in V}\left(\mathbf{M}_{w,k}^{[t_1+1,t_2]}-1/n\right)^2}\right).$$

In particular, for any node $w\in V$ and $\delta>1/n$, it holds that

$$\mathbf{Pr}\left[\left|x_w^{(t_1)}-\overline{x}\right|\geqslant\delta\right]\leqslant 2\cdot\exp\left(-\left(\delta-\frac{1}{2n}\right)^2\Big/4\right).$$

Proof. By (2.1), for any node w and round t it holds that

$$x_w^{(t)}=\xi_w^{(t)}+\sum_{s=1}^{t}\sum_{[u:v]\in\mathbf{M}^{(s)}}\left(\mathbf{M}_{u,w}^{[s+1,t]}-\mathbf{M}_{v,w}^{[s+1,t]}\right)\cdot e_{u,v}^{(s)},$$

where $\xi^{(0)}=x^{(0)}$. Therefore

$$\begin{aligned}
&\sum_{w\in V} x_w^{(t_1)}\mathbf{M}_{w,k}^{[t_1+1,t_2]}\\
&=\sum_{w\in V}\left(\xi_w^{(t_1)}+\sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}}\left(\mathbf{M}_{u,w}^{[s+1,t_1]}-\mathbf{M}_{v,w}^{[s+1,t_1]}\right)\cdot e_{u,v}^{(s)}\right)\cdot\mathbf{M}_{w,k}^{[t_1+1,t_2]}\\
&=\sum_{w\in V}\xi_w^{(t_1)}\cdot\mathbf{M}_{w,k}^{[t_1+1,t_2]}+\sum_{w\in V}\sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}}\mathbf{M}_{w,k}^{[t_1+1,t_2]}\cdot\left(\mathbf{M}_{u,w}^{[s+1,t_1]}-\mathbf{M}_{v,w}^{[s+1,t_1]}\right)\cdot e_{u,v}^{(s)}\,.
\end{aligned}$$

By Lemma B.5 and $\overline{\xi}=\overline{x}$, after t_1 rounds we have $\sum_{w\in V}\xi_w^{(t_1)}\cdot\mathbf{M}_{w,k}^{[t_1+1,t_2]}=\overline{x}\pm\frac{1}{2n}$, where equaling to $\overline{x}\pm\frac{1}{2n}$ means within the interval $\left[\overline{x}-\frac{1}{2n},\overline{x}+\frac{1}{2n}\right]$. Therefore

$$\sum_{w\in V} x_w^{(t_1)}\mathbf{M}_{w,k}^{[t_1+1,t_2]}=\overline{x}\pm\frac{1}{2n}+\sum_{w\in V}\sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}}\mathbf{M}_{w,k}^{[t_1+1,t_2]}\cdot\left(\mathbf{M}_{u,w}^{[s+1,t_1]}-\mathbf{M}_{v,w}^{[s+1,t_1]}\right)\cdot e_{u,v}^{(s)},$$

and

$$
\begin{aligned}
&\mathbf{Pr}\left[\left|\sum_{w\in V} x_w^{(t_1)}\cdot \mathbf{M}_{w,k}^{[t_1+1,t_2]} - \overline{x}\right| \geqslant \delta\right]\\
&= \mathbf{Pr}\left[\left|\left(\sum_{w\in V}\sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}} \mathbf{M}_{w,k}^{[t_1+1,t_2]}\cdot\left(\mathbf{M}_{u,w}^{[s+1,t_1]}-\mathbf{M}_{v,w}^{[s+1,t_1]}\right)\cdot e_{u,v}^{(s)}\right)\pm\frac{1}{2n}\right|\geqslant\delta\right]\\
&\leqslant \mathbf{Pr}\left[\left|\sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}}\left(\sum_{w\in V}\mathbf{M}_{w,k}^{[t_1+1,t_2]}\cdot\left(\mathbf{M}_{u,w}^{[s+1,t_1]}-\mathbf{M}_{v,w}^{[s+1,t_1]}\right)\right)\cdot e_{u,v}^{(s)}\right|\geqslant\delta-\frac{1}{2n}\right]\\
&\leqslant 2\cdot\exp\left(-\frac{(\delta-1/(2n))^2}{2\sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}}\left(\sum_{w\in V}\mathbf{M}_{w,k}^{[t_1+1,t_2]}\cdot\left(\mathbf{M}_{u,w}^{[s+1,t_1]}-\mathbf{M}_{v,w}^{[s+1,t_1]}\right)\right)^2}\right),
\end{aligned}
$$

where the last inequality follows from Lemma 2.12. Further,

$$
\begin{aligned}
\sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}}\left(\sum_{w\in V}\mathbf{M}_{w,k}^{[t_1+1,t_2]}\cdot\left(\mathbf{M}_{u,w}^{[s+1,t_1]}-\mathbf{M}_{v,w}^{[s+1,t_1]}\right)\right)^2 &= \sum_{s=1}^{t_1}\sum_{[u:v]\in\mathbf{M}^{(s)}}\left(\mathbf{M}_{u,k}^{[s+1,t_2]}-\mathbf{M}_{v,k}^{[s+1,t_2]}\right)^2\\
&\leqslant 2\cdot\sum_{w\in V}\left(\mathbf{M}_{w,k}^{[t_1+1,t_2]}-\frac{1}{n}\right)^2,
\end{aligned}
$$

where the last inequality follows from Corollary 2.11. Therefore,

$$
\mathbf{Pr}\left[\left|\sum_{w\in V} x_w^{(t_1)}\cdot\mathbf{M}_{w,k}^{[t_1+1,t_2]}-\overline{x}\right|\geqslant\delta\right]\leqslant 2\cdot\exp\left(-\frac{(\delta-1/(2n))^2}{4\sum_{w\in V}\left(\mathbf{M}_{w,k}^{[t_1+1,t_2]}-1/n\right)^2}\right),
$$

which finishes the proof of the first statement. The second statement follows directly by using the first statement with $t_1 = t_2$, since $\mathbf{M}^{[t_2+1,t_2]} = \mathbf{I}$. ■

Based on the upper bound on $\Psi_2(\mathcal{M})$, we obtain the following theorem:

Theorem 2.14. *Let G be any graph. Then, the following three statements hold:*

- *Let $\mathcal{M} = \langle \mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \ldots\rangle$ be any sequence of matchings. If $x^{(0)} = \xi^{(0)}$, then for any round t and any $\delta \geqslant 1$, it holds that*

$$
\mathbf{Pr}\left[\max_{w\in V}\left|x_w^{(t)}-\xi_w^{(t)}\right|\geqslant\sqrt{4\delta\cdot\log n}\right]\leqslant 2n^{-\delta+1}.
$$

- *In the balancing circuit model, we reach a discrepancy of $\sqrt{12\log n}+1$ after $\tau_{\text{cont}}(K,1) = \mathcal{O}\big(d\cdot\frac{\log(Kn)}{1-\lambda(\mathbf{M})}\big)$ rounds with probability at least $1-2n^{-2}$. In the random matching model, we reach a discrepancy of $\sqrt{12\log n}+1$ after $\tau_{\text{cont}}(K,1) = \mathcal{O}\big(\frac{\log(Kn)}{1-\lambda(\mathbf{P})}\big)$, with probability at least $1-2n^{-1}$.*
- *Consider the random matching model with $x^{(0)} = \xi^{(0)}$. If the initial load vector $x^{(0)}$ has discrepancy at most K, then with probability at least $1-2n^{-1}$,*

$$
\sup_{t\in\mathbb{N}}\max_{w\in V}\left|x_w^{(t)}-\xi_w^{(t)}\right|\leqslant\sqrt{4\cdot(6\log n+\log\log K)}+1.
$$

The first statement of Theorem 2.14 states that even if an adversary specifies the matchings for all rounds, it is not possible to achieve a deviation of more than $\mathcal{O}(\sqrt{\log n})$ between the discrete and the continuous case (for fixed round and fixed node).

Proof of Theorem 2.14. Recall by (2.1) that

$$x_w^{(t)} - \xi_w^{(t)} = \sum_{s=1}^{t} \sum_{[u:v]\in\mathbf{M}^{(s)}} e_{u,v}^{(s)} \left(\mathbf{M}_{u,w}^{[s+1,t]} - \mathbf{M}_{v,w}^{[s+1,t]} \right).$$

Applying Lemma 2.12 with $g_{u,v}^{(s)} = \mathbf{M}_{u,w}^{[s+1,t]} - \mathbf{M}_{v,w}^{[s+1,t]}$ yields for any $w \in V$

$$\mathbf{Pr}\left[\left| x_w^{(t)} - \xi_w^{(t)} \right| \geqslant \sqrt{2\delta \cdot \log n} \cdot \Psi_2(\mathcal{M}) \right] \leqslant 2n^{-\delta}.$$

Taking the union bound over all n nodes and recalling the bound $\Psi_2(\mathcal{M}) \leqslant \sqrt{2}$ from Theorem 2.10 completes the proof of the first statement. The second statement follows directly from by the first statement ($\delta = 3$) and the definition of $\tau_{\mathrm{cont}}(K, 1)$.

For the proof of the third statement, we require a general estimate on $\lambda_2(\mathbf{P})$. By Cheeger's inequality, $\lambda_2(\mathbf{P}) \leqslant 1 - \Phi(\mathbf{P})^2/2$, where $\Phi(\mathbf{P})$ is the conductance [36] defined as

$$\Phi(\mathbf{P}) = \min_{\substack{S\subseteq V:\\ 0<|S|\leqslant n/2}} \frac{|E(S, V \setminus S)|}{2\Delta \cdot |S|}.$$

Since G is connected, $\Phi(\mathbf{P}) \geqslant \frac{1}{n^2}$ and Cheeger's inequality implies that $\lambda_2(\mathbf{P}) \leqslant 1 - \frac{1}{2n^4}$. Using Theorem 2.5, it follows that in the continuous case, the discrepancy is at most 1 after $\tau := \mathcal{O}\left(\frac{\log(Kn)}{1-\lambda_2(\mathbf{P})}\right) = \mathcal{O}(\log(Kn) \cdot n^4)$ rounds with probability at least $1 - n^{-1}$. Using the first statement with $\delta = 6 + \log\log K/\log n$, it follows by the union bound over the time-interval $[1, \tau]$ that

$$\mathbf{Pr}\left[\max_{t\in[0,\tau]} \max_{w\in V} \left| x_w^{(t)} - \xi_w^{(t)} \right| \leqslant \sqrt{4\delta \cdot \log n} \right] \geqslant 1 - \mathcal{O}(\log(Kn) \cdot n^4) \cdot 2n^{-\delta+1} \geqslant 1 - n^{-1}.$$

Combining the three insights that (i) the the maximum load is non-increasing and the minimum load is non-decreasing, (ii) the discrepancy of the continuous case is at most 1 in round τ (w.p. at least $1 - n^{-1}$) and (iii) the maximum deviation between the discrete and continuous process is at most $\sqrt{4\delta \cdot \log n}$ in the time-interval $[0, \tau]$ (w.p. at least $1 - n^{-1}$), we conclude that the maximum deviation between the continuous and the discrete case is at most $\sqrt{4\delta \cdot \log n} + 1$ for *all* rounds. This proves the third statement and finishes the proof of the theorem. ■

3 Token-Based Analysis via Random Walks

In this section we introduce our technique which relates the movement of the tokens to independent random walks. In Section 3.1 we formalize this relation and derive strong concentration results for the load distribution on a subset of nodes. In Section 3.2 we use these new concentration results to analyze the discrepancy on arbitrary graphs. All our results in this section will hold for arbitrary for the balancing circuit and random matching model.

3.1 Bounding the Load via Random Walks

We now present our new approach that allows us to upper bound the load of a node by assuming that the tokens perform independent random walks in every round. Throughout Section 3.1, we assume that the load vector is non-negative.

Let $\mathcal{T} = \{1, \ldots, \|x^{(0)}\|_1\}$ be the set of all tokens, which are assumed to be distinguishable for the sake of the analysis. The tokens may change their location via matching edges according to the following rule: If two nodes u and v are matched in round t, then $x_u^{(t-1)} + x_v^{(t-1)}$ tokens, which are located at node u or node v at the end of round $t-1$, are placed in a single urn. After that, if $\Phi_{u,v}^{(t)} = 1$, then node u draws $\left\lceil \frac{x_u^{(t-1)} + x_v^{(t-1)}}{2} \right\rceil$ tokens from the urn uniformly at random without replacement and node v receives the remaining tokens. Otherwise, $\Phi_{u,v}^{(t)} = -1$, and node u draws $\left\lfloor \frac{x_u^{(t-1)} + x_v^{(t-1)}}{2} \right\rfloor$ tokens from the urn and again node v receives the remaining tokens. We observe that each individual token which is located at node u or v at the end of round $t-1$ is assigned to either u or v with probability $1/2$. Note that this token-based process performs exactly in the same way as the original randomized protocol introduced in Section 2.

We now prove that every token viewed individually performs a random walk with respect to the matching matrices. Henceforth we use $w_i^{(t)}$ to represent location (the node) of token $i \in \mathcal{T}$ at the end of round t. We also use the the notation that for any n by n matrix $\mathbf{M}$, any node $u \in V$ and subset $D \subseteq V$, $\mathbf{M}_{u,D} := \sum_{v \in D} \mathbf{M}_{u,D}$.

Lemma 3.1. *Fix any non-negative load vector at the end of round t_1 and consider a token $i \in \mathcal{T}$ located at node $u = w_i^{(t_1)}$ at the end of round t_1. Then for any $t_2 \geqslant t_1$,*

$$\mathbf{Pr}\left[w_i^{(t_2)} = v \right] = \mathbf{M}_{u,v}^{[t_1+1,t_2]},$$

and more generally, for any set $D \subseteq V$,

$$\mathbf{Pr}\left[w_i^{(t_2)} \in D \right] = \mathbf{M}_{u,D}^{[t_1+1,t_2]}.$$

Proof. We prove by backward induction on t that for an arbitrary pair of nodes $u, v \in V$ and round $t \in [t_1, t_2]$, the probability for a token which is at node u at the end of round t to be at node v at the end of round t_2 equals $\mathbf{M}_{u,v}^{[t+1,t_2]}$. Since $\mathbf{M}^{[t_2+1,t_2]}$ is the identity matrix, the claim is trivially true for $t = t_2$. Consider now any round $t_1 \leqslant t < t_2$ so that the induction hypothesis holds for $t+1$ and let i be a token at node u at the end of round t. If node u is not part of the matching in round $t+1$, then the induction step holds trivially. So suppose that node u is matched with a node k in round $t+1$. Since tokens are spread uniformly, it follows that the tokens at node u will be either at node u or node k with probability exactly $1/2$, regardless whether the sum of tokens at u and k is even or not. Hence,

$$\mathbf{Pr}\left[w_i^{(t_2)} = v \mid w_i^{(t)} = u \right] = \frac{1}{2} \cdot \mathbf{Pr}\left[w_i^{(t_2)} = v \mid w_i^{(t+1)} = u \right] + \frac{1}{2} \cdot \mathbf{Pr}\left[w_i^{(t_2)} = v \mid w_i^{(t+1)} = k \right]$$

Using the induction hypothesis, it follows that

$$\begin{aligned}
\mathbf{Pr}\left[w_i^{(t_2)} = v \mid w_i^{(t)} = u \right] &= \frac{1}{2} \cdot \mathbf{M}_{u,v}^{[t+2,t_2]} + \frac{1}{2} \cdot \mathbf{M}_{k,v}^{[t+2,t_2]} \\
&= \mathbf{M}_{u,u}^{(t+1)} \cdot \mathbf{M}_{u,v}^{[t+2,t_2]} + \mathbf{M}_{u,k}^{(t+1)} \cdot \mathbf{M}_{k,v}^{[t+2,t_2]} \\
&= \mathbf{M}_{u,v}^{[t+1,t_2]},
\end{aligned}$$

which completes the induction. The second statement of the lemma follows immediately by summing over all nodes in D. ■

The next lemma is the crux of our token-based analysis. It shows that the probability that a certain set of tokens will be located on a set of nodes D at the end of round t_2 is at most the product of the individual probabilities. This negative correlation will enable us to derive a strong version of the Chernoff bound (see Lemma 3.3).

Lemma 3.2. *Fix any non-negative load vector at the end of round t_1 and let $\mathcal{B} \subseteq \mathcal{T}$ be an arbitrary subset of tokens. Then for any subset of nodes $D \subseteq V$ and round $t_2 > t_1$, it holds that*

$$\mathbf{Pr}\left[\bigwedge_{i\in\mathcal{B}}\left(w_i^{(t_2)} \in D\right)\right] \leqslant \prod_{i\in\mathcal{B}} \mathbf{M}^{[t_1+1,t_2]}_{w_i^{(t_1)},D} = \prod_{i\in\mathcal{B}} \mathbf{Pr}\left[w_i^{(t_2)} \in D\right].$$

Proof. We only have to prove the inequality, as the equality follows directly from Lemma 3.1. Assume for simplicity that all tokens in $\mathcal{B}$ are numbered from 1 to $k := |\mathcal{B}|$. To simplify the notation, we define for any token $i \in \mathcal{B}$ and round t, $t_1 \leqslant t \leqslant t_2$,

$$z_i^{(t)} := \mathbf{M}^{[t+1,t_2]}_{w_i^{(t)},D}$$

and

$$Z^{(t)} := \prod_{i=1}^{k} z_i^{(t)}.$$

Our goal is to prove that the sequence $Z^{(t)}$, $t \geqslant t_1$ forms a supermartingale with respect to the sequence of load vectors $x^{(t_1)}, \ldots, x^{(t)}$, i.e.,

$$\mathbf{E}\left[\, Z^{(t)} \;\middle|\; x^{(t-1)}, \ldots, x^{(t_1)} \right] \leqslant Z^{(t-1)}. \tag{3.1}$$

Assuming that (3.1) holds, we can deduce the statement of the lemma as follows:

$$\mathbf{E}\left[\, Z^{(t_2)} \right] \leqslant \mathbf{E}\left[\, Z^{(t_1)} \right] = Z^{(t_1)} = \prod_{i=1}^{k} z_i^{(t_1)} = \prod_{i=1}^{k} \mathbf{M}^{[t_1+1,t_2]}_{w_i^{(t_1)},D}.$$

By definition,

$$Z^{(t_2)} = \prod_{i=1}^{k} \mathbf{M}^{[t_2+1,t_2]}_{w_i^{(t_2)},D},$$

which is one if $w_i^{(t_2)} \in D$ for all $i \in \mathcal{B}$ and zero otherwise. Therefore $\mathbf{E}\left[\, Z^{(t_2)} \right] = \mathbf{Pr}\left[\, Z^{(t_2)} = 1 \right] = \mathbf{Pr}\left[\bigwedge_{i\in\mathcal{B}}(w_i^{(t_2)} \in D)\right]$, and the proof is complete.

It remains to prove (3.1). To this end, fix the load vector $x^{(t-1)}$ and partition the set of tokens $\mathcal{B} = \{1, \ldots, k\}$ into disjoint sets $S_1, S_2, \ldots, S_{k'}$ with $1 \leqslant k' \leqslant k$ so that every token in S_j has the same set of possible destinations at the end of round t. Since tokens with different sets of possible destinations behave independently in round t, it follows that

$$\begin{aligned}
\mathbf{E}\left[\, Z^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)} \right] &= \mathbf{E}\left[\prod_{j=1}^{k'} \prod_{i\in S_j} z_i^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)} \right] \\
&= \prod_{j=1}^{k'} \mathbf{E}\left[\prod_{i\in S_j} z_i^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)} \right].
\end{aligned} \tag{3.2}$$

Hence in order to prove (3.1) it suffices to prove that for every fixed $j \in \{1, \ldots, k'\}$,

$$\mathbf{E}\left[\prod_{i\in S_j} z_i^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)} \right] \leqslant \prod_{i\in S_j} z_i^{(t-1)}. \tag{3.3}$$

Consider first those sets S_j so that every token $i \in S_j$ has only one destination, meaning that node $w_i^{(t-1)}$ is not incident to any matching edge in round t. In this case, clearly we have $w_i^{(t)} = w_i^{(t-1)}$ and hence $z_i^{(t)} = z_i^{(t-1)}$, and consequently (3.3) holds.

The second and more involved case concerns those sets S_j so that every token in S_j has two possible destinations at the end of round t, denoted by $u = u(j)$ and $v = v(j)$, with $\{u, v\} \in \mathbf{M}^{(t)}$. Assume for simplicity that $\mathbf{M}_{u,D}^{[t+1,t_2]} \geqslant \mathbf{M}_{v,D}^{[t+1,t_2]}$ and tokens in S_j are numbered from 1 to $\gamma = |S_j|$. Then for every token $i \in \{1, \ldots, \gamma\}$, define a random variable X_i as follows:

$$X_i = \begin{cases} 1 & \text{if token } i \text{ is assigned to node } u \text{ in round } t, \\ 0 & \text{if token } i \text{ is assigned to node } v \text{ in round } t. \end{cases}$$

Our claim is that the (random) vector $X = (X_1, \ldots, X_\gamma) \in \{0,1\}^\gamma$ satisfies the negative regression condition (cf. Definition A.4), i.e., for every two disjoint subsets $\mathcal{L}$ and $\mathcal{R}$ of S_j, and non-decreasing function $f\colon \{0,1\}^{|\mathcal{L}|} \to \mathbb{R}$, it holds that

$$\mathbf{E}\left[\, f(X_q, q \in \mathcal{L}) \;\middle|\; X_r = \sigma_r, r \in \mathcal{R} \,\right]$$

is non-increasing in each $\sigma_r \in \{0,1\}, r \in \mathcal{R}$. To establish this, it suffices to show that

$$\mathbf{E}\left[\, f(X_q, q \in \mathcal{L}) \;\middle|\; X_r = \sigma_r, r \in \mathcal{R} \,\right] \geqslant \mathbf{E}\left[\, f(X_q, q \in \mathcal{L}) \;\middle|\; X_r = \widetilde{\sigma}_r, r \in \mathcal{R} \,\right], \tag{3.4}$$

where $\widetilde{\sigma}_r = \sigma_r$ for every $r \in \mathcal{R}$ except for one $r' \in \mathcal{R}$, where $\widetilde{\sigma}_{r'} > \sigma_{r'}$. To prove the above inequality, we use a coupling argument. We expose the locations of the tokens in S_j one after another in an arbitrary order. In particular, we may expose the destinations of tokens in S_j before considering the other tokens (the ones not in $\mathcal{B}$) which are located on u and v at the beginning of round t. Note that for every token $i \in S_j$, the probability of being placed at node u (or v) depends on the placement of the previous tokens. In fact, the exact probability is not important here, instead we shall only use the fact that the probability for a token to be assigned to node u is non-increasing in the number of tokens that have been assigned to u before. For any $1 \leqslant i \leqslant \gamma + 1$, let $\alpha(i)$ be the number of tokens in $\{1, \ldots, i-1\}$ which are assigned to node u. Hence if we associate to every token $i \in S_j$ a uniform random variable $U_i \in [0,1]$, then there exists a threshold function $T(i, \alpha(i)) \in [0,1]$ satisfying the following properties:

1. if $U_i \geqslant T(i, \alpha(i))$, then token i is assigned to node u,
2. if $U_i < T(i, \alpha(i))$, then token i is assigned to node v,
3. $T(i, \alpha(i))$ is non-decreasing in $\alpha(i)$.

Without loss of generality assume that $S_j = \{1, \ldots, \gamma\}$, $\mathcal{R} = \{1, \ldots, r\}$, $\mathcal{L} = \{r+1, \ldots, r+\ell\}$, $r + \ell \leqslant \gamma$. Recall that (3.4) involves two conditional probability spaces, one for $X_r = \sigma_r, r \in \mathcal{R}$ and the other one for $X_r = \widetilde{\sigma}_r, r \in \mathcal{R}$. We denote these probability spaces by Ω and $\widetilde{\Omega}$, respectively.

Since these probability spaces are only conditional on the placements of tokens in $\mathcal{R}$, we can couple both probability spaces by assuming that the random variables U_i attain the same values for every $i \in \mathcal{L}$ in Ω and $\widetilde{\Omega}$. Further, let us denote by $\widetilde{\alpha}(i)$ the number of tokens in $\{1, \ldots, i-1\}$ which are placed on node u in $\widetilde{\Omega}$. Then, the values U_i ($i \in \mathcal{L}$), $\alpha(r+1)$ and $\widetilde{\alpha}(r+1)$ determine the placement of all tokens in $\mathcal{L}$ for the two probability spaces. By assumption on $\widetilde{\sigma}$, we have $\alpha(r+1) \in \{\widetilde{\alpha}(r+1) - 1, \widetilde{\alpha}(r+1)\}$.

Now by the three properties and the coupling described above, if for some $i \in \mathcal{L}$, $\alpha(i) = \widetilde{\alpha}(i)$, then $\alpha(i+1) = \widetilde{\alpha}(i+1)$ and all further tokens in $\mathcal{L}$ are placed in the same way in both probability spaces. Additionally, if for some $i \in \mathcal{L}$, $\alpha(i) = \widetilde{\alpha}(i) - 1$, then it follows that $\alpha(i+1) \in \{\widetilde{\alpha}(i+1) - 1, \widetilde{\alpha}(i+1)\}$ by the monotonicity property of the threshold. This means that every token $i \in \mathcal{L}$ that is placed on u in $\widetilde{\Omega}$ will be also placed on u in Ω. Since f is

non-decreasing in each coordinate (which corresponds to the placement of one token in $\mathcal{L}$), the coupling argument above establishes (3.4).

By (3.4), the vector $X = (X_1, \ldots, X_\gamma)$ satisfies the negative regression property. Then for any $i \in \{1, \ldots, \gamma\}$, define a random variable $h(X_i)$ as follows: $h(X_i) = \mathbf{M}^{[t+1,t_2]}_{u(j),D}$ if $X_i = 1$ and $h(X_i) = \mathbf{M}^{[t+1,t_2]}_{v(j),D}$ if $X_i = 0$. By the choice of u and v, we know that $h(X_i)$ is non-decreasing in every coordinate. Hence,

$$\begin{aligned}
\mathbf{E}\left[\prod_{i\in S_j} z_i^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)}\right] &= \mathbf{E}\left[\prod_{i\in S_j} h(X_i) \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)}\right] \\
&\overset{\text{Lemma A.5}}{\leqslant} \prod_{i\in S_j} \mathbf{E}\left[\, h(X_i) \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)}\right] \\
&\overset{\text{Lemma 3.1}}{=} \prod_{i\in S_j} \left(\frac{1}{2}\mathbf{M}^{[t+1,t_2]}_{u(j),D} + \frac{1}{2}\mathbf{M}^{[t+1,t_2]}_{v(j),D}\right).
\end{aligned}$$

Since $\mathbf{M}^{[t,t_2]}_{u(j),D} = \frac{1}{2}\mathbf{M}^{[t+1,t_2]}_{u(j),D} + \frac{1}{2}\mathbf{M}^{[t+1,t_2]}_{v(j),D} = \mathbf{M}^{[t,t_2]}_{v(j),D}$, we arrive at

$$\mathbf{E}\left[\prod_{i\in S_j} z_i^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)}\right] \leqslant \prod_{i\in S_j} \mathbf{M}^{[t,t_2]}_{w_i^{(t-1)},D} = \prod_{i\in S_j} z_i^{(t-1)}.$$

Applying this to (3.2) for all $j \in \{1, \ldots, k'\}$ implies that

$$\mathbf{E}\left[\, Z^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)}\right] = \prod_{j=1}^{k'} \mathbf{E}\left[\prod_{i\in S_j} z_i^{(t)} \,\middle|\, x^{(t-1)}, \ldots, x^{(t_1)}\right] \leqslant \prod_{j=1}^{k'} \prod_{i\in S_j} z_i^{(t-1)} = Z^{(t-1)},$$

showing that $Z^{(t)}$ is indeed a supermartingale. This establishes (3.1) and finishes the proof of the lemma. ■

Combining Lemma 3.2 and Lemma A.2, we obtain directly the following Chernoff bound:

Lemma 3.3. *Fix any non-negative load vector $x^{(t_1)}$ at the end of round t_1 and let $\mathcal{T}$ be the set of all tokens. Let D be any subset of nodes and $t_2 > t_1$. Then for the random variable $Z := \sum_{i\in\mathcal{T}} \mathbf{1}_{w_i^{(t_2)}\in D} = \sum_{u\in D} x_u^{(t_2)}$, it holds for any $\delta > 0$ that*

$$\mathbf{Pr}\left[\, Z \geqslant (1+\delta)\mathbf{E}\left[\, Z\,\right]\right] \leqslant \left(\frac{\mathrm{e}^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbf{E}[\,Z\,]}.$$

The strength of Lemma 3.3 is that the sum of loads is analyzed by means of a sum of indicator random variables over all tokens instead of a sum of rounding errors (e.g., Lemma 2.13). For an illustration of the power of Lemma 3.3, we consider the following toy example.

Corollary 3.4. *Let $x^{(0)}$ be any non-negative load vector with $\|x^{(0)}\|_1 \leqslant n^{1-\varepsilon}$, where $\varepsilon > 0$ is a constant. Then the discrepancy after $\tau_{\text{cont}}(1, n^{-1})$ rounds is at most $9/\varepsilon$ with probability at least $1 - 2 \cdot n^{-1}$.*

We can think of the allocation of the $\|x^{(0)}\|_1$ tokens in terms of the famous balls-and-bins model [30]. If we run our randomized protocol for sufficiently many rounds, say $\tau_{\text{cont}}(1, n^{-1})$ rounds, then every token (corresponding to a ball) is located at any node (corresponding to a bin) with almost the same probability. While in the standard balls-and-bins model, the allocation of different balls are mutually independent, Lemma 3.2 established that in our model, these allocations are negatively correlated. Therefore, as it is the case for the balls-and-bins model, we obtain a constant maximum load if the number of tokens is bounded by $n^{1-\varepsilon}$.

Proof of Corollary 3.4. Fix any node $u \in V$. By definition, it holds for $t := \tau_{\text{cont}}(1, n^{-1})$ that the time-interval $[0, t]$ is $(1, n^{-1})$–smoothing with probability at least $1 - n^{-1}$ (this probability is even 1 for the balancing circuit model). Consider $Z := \sum_{i \in \mathcal{T}} \mathbf{1}_{w_i^{(t)}}$. Assuming that $[0, t]$ is $(1, n^{-1})$–smoothing, Lemma B.5 implies that every token is located at any node u with probability at most $2/n$. Hence $\mathbf{E}[\,Z\,] \leqslant \|x^{(0)}\|_1 \cdot 2/n \leqslant 2n^{-\varepsilon} < 1$. Applying Lemma 3.3 with $\delta = \frac{8}{\varepsilon} \cdot \frac{1}{\mathbf{E}[\,Z\,]} > 9$ yields

$$\mathbf{Pr}\,[\,Z \geqslant (9/\varepsilon)\,] \leqslant \delta^{-\delta \mathbf{E}[\,Z\,]/2} \leqslant \left(\frac{\varepsilon \mathbf{E}[\,Z\,]}{8}\right)^{4/\varepsilon} \leqslant n^{-2},$$

where the inequality above holds for sufficiently large n. Taking the union bound over all nodes yields the claim. ■

The next lemma provides a concrete tail bound which is not just exponential in the deviation from the mean but also exponential in the "sparseness" of the load vector. By contrast, previous analyses expressing the load as a sum of rounding errors [8, 17, 20, 29, 35], yield weaker tail bounds for sparse load vectors (cmp. Lemma 2.13). Another advantage of Lemma 3.5 is that it gives a tail bound for an *arbitrary convex combination* of the load vector.

Lemma 3.5. *Fix any non-negative load vector $x^{(t_1)}$ with $\|x^{(t_1)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^\sigma}$ for some constant $\sigma \in (0, 1)$. Moreover, consider a round $t_2 > t_1$ so that $[t_1, t_2]$ is (n, n^{-3})–smoothing and let $Z := \sum_{v \in V} y_v x_v^{(t_2)}$, where y is any non-negative vector with $\|y\|_1 = 1$. Then for any $\delta > 0$,*

$$\mathbf{Pr}\left[\,Z \geqslant \mathrm{e}^{-\frac{1}{5}(\log n)^\sigma} + 8\|y\|_\infty \cdot (\log n)^\delta\,\right] \leqslant \mathrm{e}^{-(\log n)^{\delta+\sigma}/6}.$$

Proof. Let $\alpha := \|y\|_\infty$. Partition V into at most $2 \log_2 n$ groups defined as follows:

$$S_i := \left\{v \in V \colon 2^{-i-1} < y_v \leqslant 2^{-i}\right\}, \quad \lfloor \log_2(1/\alpha) \rfloor \leqslant i < \lceil 2 \log_2 n \rceil,$$
$$S_{\lceil 2\log_2 n \rceil} := \left\{v \in V \colon y_v \leqslant 2^{-\lceil 2 \log_2 n \rceil}\right\}.$$

Clearly, $|S_i| \leqslant 2^{i+1}$, which also holds for $S_{\lceil 2\log_2 n \rceil}$ as there are only n nodes in V. In order to obtain a bound on Z, we will upper bound the following approximation of Z:

$$\widetilde{Z} := \sum_{i = \lfloor \log_2(1/\alpha) \rfloor}^{\lceil 2 \log_2 n \rceil} \sum_{v \in S_i} x_v^{(t_2)} \cdot 2^{-i}.$$

Note that $Z \leqslant \widetilde{Z}$. We now apply our technique our relation between the movements of tokens and independent random walks to upper bound $\widetilde{Z}$. We do this by considering the contribution from each S_i individually. Since $[t_1, t_2]$ is (n, n^{-3})–smoothing, Lemma B.5 implies that every token is located at any node in round t_2 with probability at most $2/n$; thus, every token is located at a node in S_i with probability at most $2|S_i|/n$.

Suppose first that $|S_i| \geqslant \mathrm{e}^{\frac{1}{2}(\log n)^\sigma}$. By Lemma 3.2, the probability that we have more than $4|S_i| \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^\sigma}$ tokens on nodes in S_i in round t_1 is upper bounded by

$$\binom{n \cdot \mathrm{e}^{-(\log n)^\sigma}}{4|S_i| \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^\sigma}} \cdot \left(\frac{2|S_i|}{n}\right)^{4|S_i| \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^\sigma}} \leqslant \left(\frac{\mathrm{e} \cdot \mathrm{e}^{-\frac{3}{4}(\log n)^\sigma}}{2}\right)^{4|S_i| \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^\sigma}} \leqslant \mathrm{e}^{-\mathrm{e}^{\Omega((\log n)^\sigma)}}.$$

Next assume that $|S_i| \leqslant \mathrm{e}^{\frac{1}{2}(\log n)^\sigma}$. For every token $j \in \{1, \ldots, \|x^{(0)}\|_1\}$, let $X_{i,j} = 1$ if token j is located at a node in S_i in round t_1 and $X_{i,j} = 0$ otherwise. Let $X_i := \sum_{j=1}^{\|x^{(0)}\|_1} X_{i,j}$. Then,

$$\mathbf{E}[\,X_i\,] \leqslant \left\|x^{(0)}\right\|_1 \cdot \frac{2 \cdot |S_i|}{n} \leqslant \mathrm{e}^{-(\log n)^\sigma} \cdot 2\mathrm{e}^{\frac{1}{2}(\log n)^\sigma} = 2\mathrm{e}^{-\frac{1}{2}(\log n)^\sigma}.$$

Using the Chernoff bound (Lemma 3.3),

$$\mathbf{Pr}\left[\,X_i \geqslant (1+\beta)\mathbf{E}\left[\,X_i\,\right]\right] \leqslant \left(\frac{e^{\beta}}{(1+\beta)^{1+\beta}}\right)^{\mathbf{E}[\,X_i\,]} \leqslant \beta^{-\beta\cdot\mathbf{E}[\,X_i\,]/2},$$

where the second inequality holds for any $\beta \geqslant 9$. Here, we choose $\beta := (\log n)^{\delta}/\mathbf{E}\left[\,X_i\,\right] \geqslant 9$ to conclude that

$$\mathbf{Pr}\left[\,X_i \geqslant 2\cdot(\log n)^{\delta}\,\right] \leqslant (\mathbf{E}\left[\,X_i\,\right])^{(\log n)^{\delta}/2} \leqslant \left(2\mathrm{e}^{-\frac{1}{2}(\log n)^{\sigma}}\right)^{(\log n)^{\delta}/2} \leqslant \mathrm{e}^{-(\log n)^{\delta+\sigma}/5},$$

assuming n is sufficiently large.

By the union bound over at most $\lceil 2\log_2 n\rceil$ groups, we conclude that with probability at least $1-\lceil 2\log_2 n\rceil\cdot\max\left\{\mathrm{e}^{-\mathrm{e}^{\Omega((\log n)^{\sigma})}}, \mathrm{e}^{-(\log n)^{\delta+\sigma}/5}\right\} \geqslant 1-\mathrm{e}^{-(\log n)^{\delta+\sigma}/6}$,

$$\begin{aligned}
Z \leqslant \sum_{i=\lfloor\log_2(1/\alpha)\rfloor}^{\lceil 2\log_2 n\rceil}\sum_{v\in S_i} x_v^{(t_2)}\cdot 2^{-i} &= \sum_{i=\lfloor\log_2(1/\alpha)\rfloor}^{\lceil 2\log_2 n\rceil} X_i\cdot 2^{-i} \\
&\leqslant \sum_{i=\lfloor\log_2(1/\alpha)\rfloor}^{\lceil 2\log_2 n\rceil} 2^{-i}\cdot\left(\frac{4|S_i|}{\mathrm{e}^{\frac{1}{4}(\log n)^{\sigma}}}+2\cdot(\log n)^{\delta}\right) \\
&\leqslant \sum_{i=\lfloor\log_2(1/\alpha)\rfloor}^{\lceil 2\log_2 n\rceil} 2^{-i}\cdot\frac{4\cdot 2^{i+1}}{\mathrm{e}^{\frac{1}{4}(\log n)^{\sigma}}} + \sum_{i=\lfloor\log_2(1/\alpha)\rfloor}^{\lceil 2\log_2 n\rceil} 2^{-i+1}\cdot(\log n)^{\delta} \\
&\leqslant \frac{64\log_2 n}{\mathrm{e}^{\frac{1}{4}(\log n)^{\sigma}}} + 8\alpha\cdot(\log n)^{\delta} \\
&\leqslant \mathrm{e}^{-\frac{1}{5}(\log n)^{\sigma}} + 8\alpha\cdot(\log n)^{\delta}.
\end{aligned}$$

■

3.2 Bounding the Discrepancy in Arbitrary Graphs

Throughout this subsection, we assume without loss of generality that $x^{(0)} \in \mathbb{Z}^n$ is any initial load vector with $\overline{x} \in [0,1)$ (cf. Observation 2.6 for a justification). Let us fix any value $\varepsilon > 0$, not necessarily constant. Then define the following set of vectors for any $\ell \geqslant 1$:

$$\mathcal{E}_{\ell} := \left\{ x \in \mathbb{Z}^n \colon \sum_{u\in V} \max\left\{x_u - 8\ell\cdot\lceil(\log n)^{\varepsilon}\rceil - \ell, 0\right\} \leqslant 4n\cdot\mathrm{e}^{-\frac{1}{4}\cdot(\log n)^{\ell\varepsilon}}\right\}.$$

Roughly speaking, $\mathcal{E}_{\ell}$ includes all load vectors whose number of tokens above the threshold $8\ell\cdot\lceil(\log n)^{\varepsilon}\rceil+\ell$ is not too large. In particular, for any load vector $x\in\mathcal{E}_{\ell}$, $\ell \geqslant \lceil 2/\varepsilon\rceil$, the maximum load of x is at most $8\ell\cdot\lceil(\log n)^{\varepsilon}\rceil+\ell$.

The next lemma shows that if we start with a load vector in $\mathcal{E}_{\ell-1}$, then the load vector after $\tau_{\mathrm{cont}}(1,n^{-2})$ rounds will be in $\mathcal{E}_{\ell}$ with high probability.

Lemma 3.6. *For any integer $\ell \geqslant 2$, $t\in\mathbb{N}$, $\varepsilon \geqslant 16/(\log\log n)$ and any vector $x\in\mathcal{E}_{\ell-1}$,*

$$\mathbf{Pr}\left[\,x^{(t+\kappa)}\in\mathcal{E}_{\ell} \;\middle|\; x^{(t)} = x\,\right] \geqslant 1-\mathrm{e}^{-\frac{1}{4}(\log n)^{\ell\varepsilon}} - n^{-1},$$

where $\kappa := \tau_{\mathrm{cont}}(1,n^{-2})$. Furthermore, $\mathbf{Pr}\left[\,x^{(\kappa)}\in\mathcal{E}_1\,\right] \geqslant 1-\mathrm{e}^{-\frac{1}{4}(\log n)^{\varepsilon}} - 3n^{-1}$, if $\kappa := \tau_{\mathrm{cont}}(K, 1/(2n))$.

Let us briefly describe the key steps for the proof of Lemma 3.6. The proof that $x^{(\kappa)}\in\mathcal{E}_1$ (with high probability) makes use of the concentration inequality for the sum of rounding errors (Lemma 2.13, second statement), which in turn is based on our upper bound on the local 2-divergence (Theorem 2.10). The proof that, starting with a load vector in $\mathcal{E}_{\ell-1}$, we obtain a load vector which is in $\mathcal{E}_{\ell}$ after κ additional rounds is based on our new concentration inequality Lemma 3.3.

Proof of Lemma 3.6. Recall that we assume here that $\overline{x} \in [0,1)$. Let us first consider the event $x^{(\kappa)} \in \mathcal{E}_1$. Consider the following potential function in round κ:

$$\Phi^{(\kappa)} := \sum_{u \in V} \exp\left(\left(x_u^{(\kappa)} - \overline{x}\right)^2 / 16\right).$$

Since $\kappa = \tau_{\text{cont}}(K, 1/(2n))$, it follows that with probability at least $1 - n^{-1}$, the time-interval $[0, \kappa]$ is $(K, 1/(2n))$–smoothing, which we will condition on for the remainder of the proof. By the second statement of Lemma 2.13, it holds for any node $u \in V$ and any $\delta > 1/n$ that

$$\mathbf{Pr}\left[\left|x_u^{(\kappa)} - \overline{x}\right| \geqslant \delta\right] \leqslant 2\exp\left(-\left(\delta - \frac{1}{2n}\right)^2 / 4\right), \tag{3.5}$$

and therefore

$$\begin{aligned}
\mathbf{E}\left[\Phi^{(\kappa)}\right] &\leqslant n \cdot \max_{u \in V} \mathbf{E}\left[\exp\left(\left(x_u^{(\kappa)} - \overline{x}\right)^2 / 16\right)\right] \\
&= n \cdot \max_{u \in V} \sum_{k=1}^{\infty} \mathbf{Pr}\left[\exp\left(\left(x_u^{(\kappa)} - \overline{x}\right)^2 / 16\right) \geqslant k\right] \\
&\leqslant n \cdot \max_{u \in V} \left(2 + \sum_{k=3}^{\infty} \mathbf{Pr}\left[\left(x_u^{(\kappa)} - \overline{x}\right)^2 / 16 \geqslant \log k\right]\right) \\
&= n \cdot \max_{u \in V} \left(2 + \sum_{k=3}^{\infty} \mathbf{Pr}\left[\left|x_u^{(\kappa)} - \overline{x}\right| \geqslant 4\sqrt{\log k}\right]\right).
\end{aligned}$$

Combining this with (3.5), we get

$$\begin{aligned}
\mathbf{E}\left[\Phi^{(\kappa)}\right] &\leqslant n \cdot \max_{u \in V} \left(2 + \sum_{k=3}^{\infty} 2 \cdot \exp\left(-\left(4\sqrt{\log k} - \frac{1}{2n}\right)^2 / 4\right)\right) \\
&\leqslant n \cdot \left(2 + \sum_{k=3}^{\infty} 2 \cdot \mathrm{e}^{-3\log k}\right) \leqslant 4n,
\end{aligned}$$

where in the last inequality we used the fact that $\sum_{k=3}^{\infty} 2k^{-3} \leqslant \frac{2}{3}\sum_{k=3}^{\infty} k^{-2} \leqslant 2$. Hence by Markov's inequality,

$$\mathbf{Pr}\left[\Phi^{(\kappa)} \geqslant 4n \cdot \mathrm{e}^{\frac{1}{4}\cdot(\log n)^{\varepsilon}}\right] \leqslant \mathbf{Pr}\left[\Phi^{(\kappa)} \geqslant \mathrm{e}^{\frac{1}{4}\cdot(\log n)^{\varepsilon}} \cdot \mathbf{E}\left[\Phi^{(\kappa)}\right]\right] \leqslant \mathrm{e}^{-\frac{1}{4}\cdot(\log n)^{\varepsilon}}. \tag{3.6}$$

Furthermore, recall that by Theorem 2.14, the maximum load at the end of round κ is upper bounded by $\sqrt{12\log n} + 2$ with probability at least $1 - 2n^{-1}$. In the following we tacitly assume that both $\Phi^{(\kappa)} \leqslant 4n \cdot \mathrm{e}^{\frac{1}{4}\cdot(\log n)^{\varepsilon}}$ and $x_{\max}^{(\kappa)} \leqslant \sqrt{12\log n} + 2$ hold. Under this condition, the total number of tokens above the threshold $8 \cdot \lceil (\log n)^{\varepsilon} \rceil + 1$ in round κ is upper bounded by

$$\sum_{u \in V} \max\left\{x_u^{(\kappa)} - 8 \cdot \lceil (\log n)^{\varepsilon} \rceil - 1, 0\right\} \leqslant \frac{4n \cdot \mathrm{e}^{\frac{1}{4}\cdot(\log n)^{\varepsilon}}}{\mathrm{e}^{4\cdot(\log n)^{2\varepsilon}}} \cdot \left(\sqrt{12\log n} + 2\right) \leqslant 4n \cdot \mathrm{e}^{-(\log n)^{\varepsilon}}, \tag{3.7}$$

where the second last inequality is due to the fact that

$$\exp\left((8 \cdot \lceil (\log n)^{\varepsilon} \rceil + 1 - \overline{x})^2 / 16\right) \geqslant \mathrm{e}^{4\cdot(\log n)^{2\varepsilon}}.$$

Combining (3.6) and (3.7) we have

$$\mathbf{Pr}\left[\sum_{u \in V} \max\left\{x_u^{(\kappa)} - 8 \cdot \lceil (\log n)^{\varepsilon} \rceil - 1, 0\right\} \geqslant 4n \cdot \mathrm{e}^{-(\log n)^{\varepsilon}}\right] \leqslant \mathrm{e}^{-\frac{1}{4}\cdot(\log n)^{\varepsilon}},$$

which implies

$$\mathbf{Pr}\left[x^{(\kappa)} \in \mathcal{E}_1 \right] \geqslant 1 - \mathrm{e}^{-\frac{1}{4}\cdot(\log n)^{\varepsilon}},$$

completing the proof of the first statement.

For the second statement with $\ell \geqslant 2$, we consider the probability space conditioned on $x^{(t)} = x \in \mathcal{E}_{\ell-1}$. To simplify the notation, we will ignore this condition in the following probabilities and expectations. For any round $s \geqslant t$, define another load vector $\widetilde{x}^{(s)}$ as follows:

$$\widetilde{x}_u^{(s)} := \max\left\{ x_u^{(s)} - 8(\ell-1)\cdot\lceil(\log n)^{\varepsilon}\rceil - (\ell-1), 0 \right\}.$$

Due to Observation 2.6, we can use $\widetilde{x}_u^{(s)}$ to upper bound the load of $x_u^{(s)}$ by using the inequality $x_u^{(s)} \leqslant \widetilde{x}_u^{(s)} + 8(\ell-1)\cdot\lceil(\log n)^{\varepsilon}\rceil + (\ell-1)$, valid for any $u \in V$ and $s \geqslant t$. Therefore, we focus on the (non-negative) load vector $\widetilde{x}^{(s)}$ in the following. Fix an arbitrary node $u \in V$. Let Z_i be the 0/1-indicator random variable for every token $i \in \mathcal{T}$, which is one if and only if token i reaches node u at the end of round $t+\kappa$. Let $\beta := \|\widetilde{x}^{(t)}\|_1$ and $Z := \sum_{i=1}^{\beta} Z_i$. Clearly, $\widetilde{x}_u^{(t+\kappa)} = Z$.

Note that for token i which is located at node $w_i^{(t)}$ at the end of round t, we have by Lemma 3.1 that

$$\mathbf{Pr}\left[Z_i = 1 \right] = \mathbf{M}^{[t+1,t+\kappa]}_{w_i^{(t)},u}.$$

Since $\kappa = \tau_{\mathrm{cont}}(1, n^{-2})$, the time-interval $[t, t+\kappa]$ is $(1, n^{-2})$–smoothing with probability at least $1 - n^{-1}$, which we will assume in the following. Hence Lemma B.5 yields for every pair of nodes $v, u \in V$,

$$\mathbf{M}^{[t+1,t+\kappa]}_{v,u} \leqslant \frac{1}{n} + \frac{1}{n^2}.$$

By Lemma 3.3, we can upper bound Z as follows: For any $\delta > 0$,

$$\mathbf{Pr}\left[Z \geqslant (1+\delta)\mathbf{E}\left[Z \right] \right] \leqslant \left(\frac{\mathrm{e}^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mathbf{E}[Z]} \leqslant \left(\frac{\mathrm{e}}{\delta} \right)^{\delta\cdot\mathbf{E}[Z]}. \tag{3.8}$$

By definition of the set $\mathcal{E}_{\ell-1}$,

$$\beta = \left\| \widetilde{x}^{(t)} \right\|_1 = \sum_{u\in V} \max\left\{ x_u^{(t)} - 8(\ell-1)\cdot\lceil(\log n)^{\varepsilon}\rceil - (\ell-1), 0 \right\} \leqslant 4n \cdot \mathrm{e}^{-\frac{1}{4}\cdot(\log n)^{(\ell-1)\cdot\varepsilon}}.$$

Hence, for sufficiently large n,

$$\mathbf{E}\left[Z \right] \leqslant \sum_{i=1}^{\beta} \left(\frac{1}{n} + \frac{1}{n^2} \right) \leqslant 5\mathrm{e}^{-\frac{1}{4}\cdot(\log n)^{(\ell-1)\cdot\varepsilon}}. \tag{3.9}$$

Since $\mathbf{E}\left[Z \right] \leqslant 1$,

$$\mathbf{Pr}\left[\widetilde{x}_u^{(t+\kappa)} \geqslant \delta\,\mathbf{E}\left[Z \right] + 1 \right] = \mathbf{Pr}\left[Z \geqslant \delta\,\mathbf{E}\left[Z \right] + 1 \right] \leqslant \mathbf{Pr}\left[Z \geqslant (1+\delta)\mathbf{E}\left[Z \right] \right].$$

Choosing $\delta = \delta(\alpha) = \frac{1}{\mathbf{E}[Z]}\cdot(8\cdot(\log n)^{\varepsilon} + \alpha)$ (for any integer $\alpha \geqslant 0$) in (3.8), we obtain

$$\begin{aligned}
\mathbf{Pr}\left[\widetilde{x}_u^{(t+\kappa)} \geqslant (8\cdot(\log n)^{\varepsilon} + \alpha) + 1 \right] &\leqslant \left(\frac{\delta}{\mathrm{e}} \right)^{-(8\cdot(\log n)^{\varepsilon}+\alpha)} \\
&\leqslant \left(\frac{5}{\mathbf{E}\left[Z \right]} \right)^{-(8\cdot(\log n)^{\varepsilon}+\alpha)} \\
&\overset{\text{by (3.9)}}{\leqslant} \exp\left(-\frac{1}{4}\cdot(\log n)^{(\ell-1)\cdot\varepsilon}\cdot(8\cdot(\log n)^{\varepsilon}+\alpha) \right), \qquad (3.10)
\end{aligned}$$

where in the second inequality we used that by our lower bound on ε, $8 \cdot (\log n)^{\varepsilon} \geqslant 5$. Our goal is now to bound the number of tokens in the load vector $\widetilde{x}^{(t+\kappa)}$ that are above the threshold $8 \cdot \lceil (\log n)^{\varepsilon} \rceil + 1$. To this end, define the potential function $\Lambda^{(t+\kappa)}$ with respect to load vector $\widetilde{x}^{(t+k)}$ by

$$\Lambda^{(t+\kappa)} := \sum_{u \in V} \max \left\{ \widetilde{x}_u^{(t+\kappa)} - 8 \cdot \lceil (\log n)^{\varepsilon} \rceil - 1, 0 \right\}.$$

Then we can upper bound the expectation of $\Lambda^{(t+\kappa)}$ as follows:

$$\begin{aligned}
\mathbf{E}\left[\Lambda^{(t+\kappa)} \right] &= \sum_{u \in V} \sum_{\alpha=1}^{\infty} \mathbf{Pr}\left[\max \left\{ \widetilde{x}_u^{(t+\kappa)} - 8 \cdot \lceil (\log n)^{\varepsilon} \rceil - 1, 0 \right\} \geqslant \alpha \right] \\
&= \sum_{u \in V} \sum_{\alpha=1}^{\infty} \mathbf{Pr}\left[\widetilde{x}_u^{(t+\kappa)} \geqslant 8 \cdot \lceil (\log n)^{\varepsilon} \rceil + 1 + \alpha \right] \\
&\overset{\text{by (3.10)}}{\leqslant} \sum_{u \in V} \sum_{\alpha=1}^{\infty} \exp\left(-\frac{1}{4} (\log n)^{(\ell-1)\cdot\varepsilon} \cdot (8 \cdot (\log n)^{\varepsilon} + \alpha) \right) .
\end{aligned}$$

Since $\frac{1}{4} \cdot (\log n)^{(\ell-1)\cdot\varepsilon} \geqslant 1$ we have

$$\begin{aligned}
\mathbf{E}\left[\Lambda^{(t+\kappa)} \right] &\leqslant n \cdot \exp\left(-(\log n)^{(\ell-1)\cdot\varepsilon+\varepsilon} \right) \cdot \sum_{\alpha=1}^{\infty} \mathrm{e}^{-\alpha} \\
&\leqslant n \cdot \mathrm{e}^{-(\log n)^{\ell\cdot\varepsilon}} \cdot \frac{1}{1-\mathrm{e}^{-1}} \leqslant 4n \cdot \mathrm{e}^{-(\log n)^{\ell\cdot\varepsilon}}.
\end{aligned}$$

Using Markov's inequality,

$$\mathbf{Pr}\left[\Lambda^{(t+\kappa)} \geqslant 4n \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^{\ell\cdot\varepsilon}} \right] \leqslant \mathrm{e}^{-\frac{3}{4}(\log n)^{\ell\cdot\varepsilon}}.$$

Assuming that $\Lambda^{(t+\kappa)} \leqslant 4n \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^{\ell\cdot\varepsilon}}$ occurs, it follows by the definition of $\Lambda^{(t+\kappa)}$ that

$$\Lambda^{(t+\kappa)} = \sum_{u \in V} \max \left\{ \widetilde{x}_u^{(t+\kappa)} - 8 \cdot \lceil (\log n)^{\varepsilon} \rceil - 1, 0 \right\} \leqslant 4n \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^{\ell\cdot\varepsilon}}.$$

Since $x_u^{(t+\kappa)} \leqslant \widetilde{x}_u^{(t+\kappa)} + 8(\ell-1) \cdot \lceil (\log n)^{\varepsilon} \rceil + (\ell - 1)$,

$$\sum_{u \in V} \max \left\{ x_u^{(t+\kappa)} - 8\ell \cdot \lceil (\log n)^{\varepsilon} \rceil - \ell, 0 \right\} \leqslant 4n \cdot \mathrm{e}^{-\frac{1}{4}(\log n)^{\ell\cdot\varepsilon}}.$$

Therefore,

$$\mathbf{Pr}\left[x^{(t+\kappa)} \in \mathcal{E}_{\ell} \right] \geqslant 1 - \mathrm{e}^{-\frac{3}{4}(\log n)^{\ell\cdot\varepsilon}} \geqslant 1 - \mathrm{e}^{-\frac{1}{4}(\log n)^{\ell\cdot\varepsilon}},$$

which finishes the proof. ■

Iterating Lemma 3.6 reveals an interesting tradeoff. First, we obtain a discrepancy of $\mathcal{O}((\log n)^{\varepsilon})$ for an arbitrarily small constant $\varepsilon > 0$ by increasing the runtime $\tau_{\text{cont}}(K, n^{-2})$ only by a constant factor. Furthermore, choosing $\varepsilon = \Theta(1/(\log \log n))$ and ℓ appropriately, we obtain a discrepancy of $\mathcal{O}(\log \log n)$ by increasing the runtime by a factor of $\mathcal{O}(\log \log n)$.

Theorem 3.7. *Let G be any graph and consider the random matching or balancing circuit model.*

- *Let $\varepsilon > 0$ be an arbitrarily small constant. Then after $\mathcal{O}(\tau_{\text{cont}}(K, n^{-2}))$ rounds, the discrepancy is $\mathcal{O}((\log n)^{\varepsilon})$ w. p. $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$.*

- *After* $\mathcal{O}(\tau_{\text{cont}}(K, n^{-2}) \cdot \log\log n)$ *rounds, the discrepancy is* $\mathcal{O}(\log\log n)$ *w. p.* $1 - \frac{1}{\log n}$.

Proof. By Lemma 3.6, with $\kappa := \tau_{\text{cont}}(K, 1/(2n))$, for any vector $x \in \mathcal{E}_{\ell-1}$ and any round $t \in \mathbb{N}$,

$$\begin{aligned}\mathbf{Pr}\left[x^{(t+\kappa)} \in \mathcal{E}_\ell \mid x^{(t)} = x\right] &\geqslant 1 - \mathrm{e}^{-\frac{1}{4}(\log n)^{\ell\cdot\varepsilon}} - 3n^{-1} \\ &\geqslant 1 - \mathrm{e}^{-\frac{1}{4}(\log n)^{\varepsilon}} - 3n^{-1} =: p, \end{aligned} \tag{3.11}$$

and the same lower bound also holds for $\mathbf{Pr}\left[x^{(\kappa)} \in \mathcal{E}_1\right]$. Our goal is to show that $x^{(\ell\cdot\kappa)}$ is in $\mathcal{E}_\ell$, where $\ell := \lceil \frac{2}{\varepsilon} \rceil$. Applying (3.11) ℓ-times and the union bound,

$$\mathbf{Pr}\left[x^{(\ell\cdot\kappa)} \in \mathcal{E}_\ell\right] \geqslant 1 - \ell \cdot \left(\mathrm{e}^{-\frac{1}{4}(\log n)^{\varepsilon}} + 3n^{-1}\right) \geqslant 1 - \mathrm{e}^{-\frac{1}{5}(\log n)^{\varepsilon}},$$

where the second inequality holds since ε and ℓ are constants. If the load vector is in $\mathcal{E}_\ell$, then

$$\sum_{u\in V} \max\left\{x_u^{(\ell\cdot\kappa)} - 8\left\lceil\frac{2}{\varepsilon}\right\rceil \cdot \lceil(\log n)^{\varepsilon}\rceil - \left\lceil\frac{2}{\varepsilon}\right\rceil, 0\right\} \leqslant 4n \cdot \exp\left(-\frac{1}{4}\cdot(\log n)^{\lceil\frac{2}{\varepsilon}\rceil\cdot\varepsilon}\right) < 1,$$

which implies that the maximum load in round $\ell\cdot\tau_{\text{cont}}(1, n^{-2})$ is $\mathcal{O}((\log n)^{\varepsilon})$. The corresponding lower bound on the minimum load follows by symmetry (see Lemma 2.7). Hence with probability $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$, the discrepancy in round $\ell \cdot \tau_{\text{cont}}(1, n^{-2})$ is $\mathcal{O}((\log n)^{\varepsilon})$, completing the proof of the first statement.

Let us now prove the second statement. First observe that if $x^{(t)} \in \mathcal{E}_\ell$ for some round t, then also $x^{(t+1)} \in \mathcal{E}_\ell$. We now choose $\varepsilon := \frac{16}{\log\log n}$, $\ell := \lceil\log\log n\rceil$ and bound the number of rounds required to reach a load vector which is in $\mathcal{E}_\ell$. We divide this time into phases each of which is of length $\max\{\tau_{\text{cont}}(K, 1/(2n)), \tau_{\text{cont}}(1, n^{-2})\} \leqslant \tau_{\text{cont}}(K, n^{-2}) =: \tau$. As the success probability p in (3.11) is only a positive constant for our choice of ε, we have to repeat some of the phases. However, the number of repetitions R before we reach a load vector in $\mathcal{E}_\ell$ is stochastically smaller than the sum of ℓ independent geometric random variables each of which has success probability p. Hence by Lemma A.3, with probability $1 - \exp(-\Omega(\log\log n)) \geqslant 1 - \frac{1}{2\cdot\log n}$, it holds that $R = \mathcal{O}(\ell)$, i.e., the load vector is in $\mathcal{E}_\ell$ after at most $\mathcal{O}(\ell)$ repetitions of τ many rounds. If the load vector is in $\mathcal{E}_\ell$, then

$$\sum_{u\in V} \max\left\{x_u^{(R\cdot\tau)} - 8\ell \cdot \lceil(\log n)^{\varepsilon}\rceil - \ell, 0\right\} \leqslant 4n \cdot \mathrm{e}^{-\frac{1}{4}\cdot(\log n)^{\ell\varepsilon}}.$$

Plugging in the values of ε and ℓ yields

$$\begin{aligned}&\sum_{u\in V} \max\left\{x_u^{(R\cdot\tau)} - 8\lceil\log\log n\rceil \cdot \left\lceil(\log n)^{\frac{16}{\log\log n}}\right\rceil - \lceil\log\log n\rceil, 0\right\} \\ &\qquad \leqslant 4n \cdot \exp\left(-\frac{1}{4}\cdot(\log n)^{\lceil\log\log n\rceil\cdot\frac{16}{\log\log n}}\right) < 1,\end{aligned}$$

where the last inequality holds for sufficiently large n. But from the last inequality it follows directly that for all nodes $u \in V$,

$$x_u^{(R\cdot\tau)} \leqslant 8\lceil\log\log n\rceil \cdot \lceil\mathrm{e}^{16}\rceil + \lceil\log\log n\rceil.$$

To get the corresponding lower bound for the minimum load, we use again Lemma 2.7. Thus with probability at least $1-\frac{1}{\log n}$, the discrepancy in round $R\cdot\tau$ is upper bounded by $\mathcal{O}(\log\log n)$. This finishes the proof of the second statement and the proof of the theorem. ■

4 Proof of the Main Theorem (Theorem 1.1)

In the remainder of this paper we sketch the proof of Theorem 1.1. For the ease of the analysis we "subtract" the same number of tokens from every node such that the resulting load vector x satisfies $\overline{x} \in [0, 1)$ (cmp. discussion after eq. (2.1)). As illustrated in Figure 1, our proof consists of the following three main steps.

1. **Reducing the Discrepancy to $(\log n)^{\varepsilon_d}$.** We first use Theorem 3.7 from Section 3 to show that in round $t_1 := \mathcal{O}\left(\tau_{\text{cont}}(K, n^{-2})\right) = \mathcal{O}\big(\frac{\log(Kn)}{1-\lambda}\big)$ the discrepancy is at most $(\log n)^{\varepsilon_d}$, where $\varepsilon_d > 0$ is an arbitrarily small constant.
2. **Sparsification of the Load Vector.** Since our goal is to achieve a constant discrepancy, we fix a constant $C > 0$ and only consider nodes with more than C tokens. We prove in Theorem 4.2 that the number of tokens above the threshold C on these nodes is at most $n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$ in round $t_2 := t_1 + \mathcal{O}\big(\frac{\log n}{1-\lambda}\big)$. The proof of this step is based on a polynomial potential function and exploits that the load vector in round t_1 has small discrepancy.
3. **Reducing the Discrepancy to a Constant.** Now we only need to analyze the $n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$ tokens above the threshold C. This is equivalent to analyze a non-negative load vector with at most $n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$ tokens (Observation 2.6). We prove in Theorem 4.3 that in round $t_3 := t_2 + \mathcal{O}\big(\frac{\log n}{1-\lambda}\big)$, there is no token above the threshold $C + 1$, using the token-based analysis via random walks (Section 3). This upper bounds the maximum load; the lower bound on the minimum load follows by symmetry (Lemma 2.7). These two bounds together imply that the discrepancy in round t_3 is at most $2C + 2$.

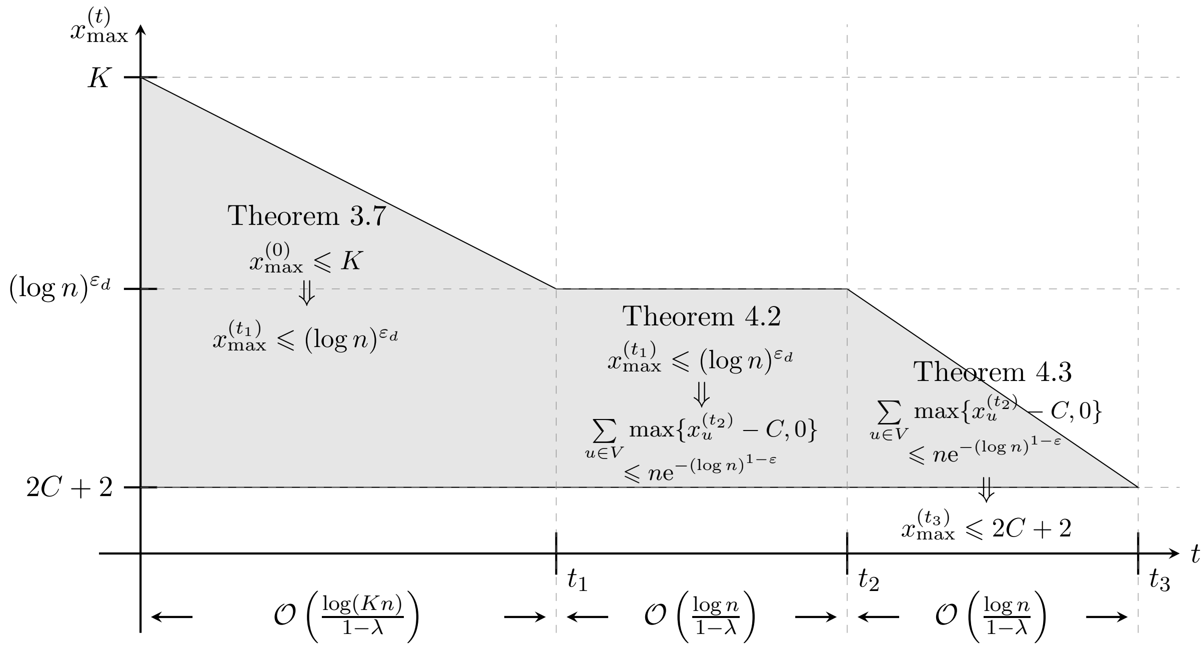


Figure 1: The above diagram illustrates how Theorem 3.7, Theorem 4.2 and Theorem 4.3 are combined to prove Theorem 1.1. We assume w.l.o.g. that $\overline{x} \in [0, 1)$ and consider only the drop of the maximum load (cf. Lemma 2.7 for the relation between upper bounding the maximum load and lower bounding the minimum load).

Remark 4.1. *All results and arguments in this section will hold for the balancing circuit model (with a constant d) and the random matching model as described in Section 2 unless mentioned otherwise. In the analysis, one round in the random matching model corresponds to d consecutive rounds in the balancing circuit model, which ensures smooth convergence as we periodically apply the same sequence of d matchings. In fact, many of the complications in the proof come from*

the random matching model, as some nodes may not be part of the randomly chosen matchings for a long period of rounds.

4.1 Proof of Theorem 1.1

In this subsection we state Theorem 4.2 and Theorem 4.3 with their proofs deferred to Section 4.2 and Section 4.3, respectively. By assuming the correctness of these two theorems, we prove our main result, Theorem 1.1, at the end of this subsection.

Theorem 4.2. *Let G be any regular graph and $\varepsilon > 0$ be any constant. Then there are constants $\varepsilon_d = \varepsilon_d(\varepsilon) > 0$ and $C = C(\varepsilon) > 0$ such that the following holds. Any load vector $x^{(0)}$ with discrepancy at most $(\log n)^{\varepsilon_d}$ and $\overline{x} \in [0, 1)$, satisfies with probability $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$ the following inequality in round $\tau := \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$:*

$$\sum_{u \in V} \max\left\{x_u^{(\tau)} - C, 0\right\} \leqslant n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}.$$

The basic idea to prove Theorem 4.2 is to consider the following potential function $\Phi^{(t)} = \sum_{u \in V \colon x_u^{(t)} \geqslant 11} \left(x_u^{(t)}\right)^8$ and show that it drops significantly. The complete proof of Theorem 4.2 is given in Section 4.2.

Theorem 4.3. *Let G be any d-regular graph. Let $\varepsilon > 0$ be a sufficiently small constant and assume that $x^{(0)}$ is a non-negative load vector with $\|x^{(0)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$. Then with probability at least $1 - 5n^{-1}$, it holds after $\kappa := \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$ rounds that $\|x^{(\kappa)}\|_\infty \leqslant 1$.*

To show Theorem 4.3, we proceed similarly as in the proof of Theorem 4.2. However, here we employ an exponential potential function that runs over all nodes with load larger than 1. Exploiting the sparseness of $x^{(0)}$, we show that after $\mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$ rounds, the value of the potential is at most n^2. Then we exploit the sparseness again to derive an upper bound on the collision probability of two tokens. From this we conclude that after β rounds, the potential is reduced by a factor of $\mathrm{e}^{\Omega(\beta \cdot (1-\lambda))}$, meaning that on average the potential function drops exponentially every $\mathcal{O}\left(\frac{1}{1-\lambda}\right)$ rounds.

We defer the proof of Theorem 4.3 to Section 4.3 and first prove Theorem 1.1 assuming the correctness of Theorem 4.2 and Theorem 4.3.

Proof of Theorem 1.1. Let $\varepsilon > 0$ be the small constant required for Theorem 4.3, which in turns gives us a constant $\varepsilon_d = \varepsilon_d(\varepsilon) > 0$ required for Theorem 4.2. By Theorem 3.7, the discrepancy is at most $(\log n)^{\varepsilon_d}$ with probability at least $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$ in round $t_1 := \mathcal{O}\big(\frac{\log(Kn)}{1-\lambda}\big)$. Next we apply Theorem 4.2 to prove that with probability at least $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$, the load vector $x^{(t_2)}$ in round $t_2 := t_1 + \mathcal{O}\big(\frac{\log n}{1-\lambda}\big)$ satisfies

$$\sum_{w \in V} \max\left\{x_w^{(t_2)} - C, 0\right\} \leqslant n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}.$$

Next define for any round $s \geqslant t_2$ a new load vector $\widetilde{x}^{(s)}$ by $\widetilde{x}_u^{(s)} := \max\big\{x_u^{(s)} - C, 0\big\}$ for any $u \in V$. Since by Observation 2.6, $x_u^{(s)} \leqslant \widetilde{x}_u^{(s)} + C$ for every $s \geqslant t_2$, it suffices to bound the maximum load of the non-negative load vector $\widetilde{x}^{(s)}$ for an upper bound on the maximum load of $x^{(s)}$. Since $\|\widetilde{x}^{(s)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$, we apply Theorem 4.3 to conclude in round $t_3 := t_2 + \mathcal{O}\big(\frac{\log n}{1-\lambda}\big)$, $\|\widetilde{x}^{(t_3)}\|_\infty \leqslant 1$ holds with probability at least $1 - 5n^{-1}$. Hence by the union bound and the relation between $\widetilde{x}^{(t_3)}$ and $x^{(t_3)}$, the maximum load of $x^{(t_3)}$ is at most $C + 1$ with probability at least $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$. The corresponding lower bound on the minimum load is derived by symmetry (see Lemma 2.7). ■

4.2 Proof of Theorem 4.2

Throughout the proof of Theorem 4.2, we use the following potential function:

$$\Phi^{(t)} := \sum_{u \in V \,:\, x_u^{(t)} \geqslant 11} \left(x_u^{(t)}\right)^8. \tag{4.1}$$

Occasionally, we will also apply this potential function to a different sequence of load vectors $\widetilde{x}^{(t)}, t \geqslant 0$, and denote this by $\Phi^{(t)}(\widetilde{x})$. Our next observation is that $\Phi^{(t)}$ is non-increasing in t. Since our protocol only transfers tokens from nodes with larger load to other nodes with smaller load, it suffices to show that $x \mapsto x^8 \cdot \mathbf{1}_{x \geqslant 11}$ is convex, which follows from the convexity of $x \mapsto x^8$ and $11^8 - 0 \leqslant 12^8 - 11^8$.

The key step in proving Theorem 4.2 is to analyze the drop of the potential function Φ, which is formalized in the following lemma.

Lemma 4.4. *Fix a constant $\sigma \in (0,1)$ and let $\tau := \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$. Then for sufficiently small constant $\varepsilon_d \in (0,1)$ (independent of σ) the following two statements hold.*

- *For any load vector $x^{(t)}$ at the end of round t with discrepancy at most $(\log n)^{\varepsilon_d}$, it holds with probability $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$ that*

$$\Phi^{(t+\tau)} \leqslant n \cdot \mathrm{e}^{-(\log n)^{\frac{1}{24}}}.$$

- *If the load vector $x^{(t)}$ is non-negative, has discrepancy at most $(\log n)^{\varepsilon_d}$ and satisfies $\|x^{(t)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{\sigma}}$, then it holds with probability $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$ that*

$$\Phi^{(t+\tau)} \leqslant n \cdot \exp\left(-(\log n)^{1-11\varepsilon_d - \frac{38}{39}(1-11\varepsilon_d-\sigma)}\right).$$

Since $\sum_{u \in V} \max\left\{x_u^{(t+\tau)} - 10, 0\right\} \leqslant \Phi^{(t+\tau)}$, the second statement of Lemma 4.4 states that the number of tokens above the threshold 10 at the end of round $t+\tau$ is much smaller than the total number of tokens in round t (assuming the total number tokens is less than $n \cdot \mathrm{e}^{-(\log n)^{\sigma}}$, which can be achieved by a single application of the first statement of Lemma 4.4). This argument can be iterated a constant number of times to obtain that the number of tokens above the threshold $10 \cdot k$ at the end of round $t + k \cdot \tau$ is at most $n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$ for a sufficiently large constant k, yielding Theorem 4.2.

4.2.1 Proof of Theorem 4.2 using Lemma 4.4.

We now proceed with the formal proof of Theorem 4.2 assuming the correctness of Lemma 4.4, whose proof is given in Section 4.2.2.

Proof of Theorem 4.2. By assumption, the discrepancy of the load vector $x^{(t)}$ is at most $(\log n)^{\varepsilon_d}$, where $t := \mathcal{O}(\tau_{\mathrm{cont}}(K, n^{-3}))$ and $\varepsilon_d \leqslant \varepsilon/22$ small enough (depending on the requirement of Lemma 4.4). Then the first statement of Lemma 4.4 implies that with probability $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$,

$$\sum_{u \in V} \max\left\{x_u^{(t+\tau)} - 10, 0\right\} \leqslant \Phi^{(t+\tau)} \leqslant n \cdot \mathrm{e}^{-(\log n)^{1/24}}.$$

Let us now define the load vector $\widetilde{x}^{(s)} := \max\left\{x_u^{(s)} - 10, 0\right\}$ for any $s \geqslant t+\tau$. By Observation 2.6, it holds for any $s \geqslant t + \tau$ and node $u \in V$ that

$$x_u^{(s)} \leqslant \widetilde{x}_u^{(s)} + 10,$$

which allows us to work with the non-negative load vector $\widetilde{x}$ in the following. By definition of $\widetilde{x}$ and the condition on $x^{(t+\tau)}$, $\left\|\widetilde{x}^{(t+\tau)}\right\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{1/24}}$. Applying the second statement of Lemma 4.4, it follows with probability $1 - \mathrm{e}^{-(\log n)^{\Omega(1)}}$ that

$$\sum_{u \in V} \max\left\{\widetilde{x}_u^{(t+2\tau)} - 10, 0\right\} \leqslant \Phi^{(t+2\tau)}(\widetilde{x}) \leqslant n \cdot \exp\left(-(\log n)^{1-11\varepsilon_d - \frac{38}{39}(1-11\varepsilon_d - a(1))}\right),$$

where $a(1) := \frac{1}{24}$. Consequently,

$$\begin{aligned}\sum_{u \in V} \max\left\{x_u^{(t+2\tau)} - 2\cdot 10, 0\right\} &\leqslant \sum_{u \in V} \max\left\{\widetilde{x}_u^{(t+2\tau)} - 10, 0\right\} \\ &\leqslant n \cdot \exp\left(-(\log n)^{1-11\varepsilon_d - \frac{38}{39}(1-11\varepsilon_d - a(1))}\right).\end{aligned}$$

Since the sequence $a(i), i \in \mathbb{N}$ defined by the recursion

$$a(i) := (1 - 11\varepsilon_d) - \frac{38}{39}\Big((1 - 11\varepsilon_d) - a(i-1)\Big),$$

$a(1) = \frac{1}{24}$, is non-decreasing and converges to $1 - 11\varepsilon_d$, it follows by the union bound that with probability at least $1 - k \cdot \mathrm{e}^{-(\log n)^{\Omega(1)}}$, it holds for any integer $k \in \mathbb{N}$ that

$$\sum_{u \in V} \max\left\{x_u^{(t+k\tau)} - k \cdot 10, 0\right\} \leqslant \Phi^{(t+k\tau)} \leqslant n \cdot \mathrm{e}^{-(\log n)^{a(k)}}.$$

Further, for any $\varepsilon > 0$, there exists a (large) constant $C = C(\varepsilon) > 0$ so that for any $k \geqslant C$, $a(k) \geqslant 1 - 11\varepsilon_d - \varepsilon/2$. Since $\varepsilon_d \leqslant \varepsilon/22$, Theorem 4.2 follows. ■

4.2.2 Proof of Lemma 4.4

This part is devoted to the proof of Lemma 4.4. First, we define canonical paths.

Definition 4.5 ([17])**.** *The sequence $\mathcal{P}_v = (\mathcal{P}_v^{(t_1)} = v, \mathcal{P}_v^{(t_1+1)}, \ldots)$ is called the* canonical path *of v from round t_1 if for all rounds t with $t > t_1$ the following holds. If $v_t := \mathcal{P}_v^{(t)}$ is unmatched in $\mathbf{M}^{(t+1)}$, then $v_{t+1} = v_t$ and $\mathcal{P}_v^{(t+1)} := v_{t+1}$. Otherwise, let $u \in V$ be the node such that $\{v_t, u\} \in \mathbf{M}^{(t+1)}$. Then,*

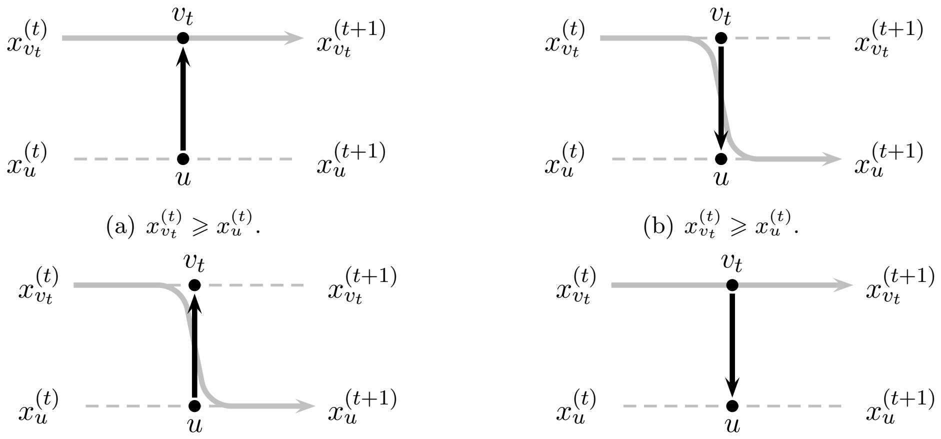
- *if $x_{v_t}^{(t)} \geqslant x_u^{(t)}$ and $\Phi_{v_t,u}^{(t+1)} = 1$ then $v_{t+1} = v_t$,*
- *if $x_{v_t}^{(t)} \geqslant x_u^{(t)}$ and $\Phi_{v_t,u}^{(t+1)} = -1$ then $v_{t+1} = u$,*
- *if $x_{v_t}^{(t)} < x_u^{(t)}$ and $\Phi_{v_t,u}^{(t+1)} = 1$ then $v_{t+1} = u$,*
- *if $x_{v_t}^{(t)} < x_u^{(t)}$ and $\Phi_{v_t,u}^{(t+1)} = -1$ then $v_{t+1} = v_t$.*

An illustration of this definition is given in Figure 2.

Note that there are always exactly n canonical paths, which are all vertex-disjoint. We define canonical paths so that if two of them are connected by a matching edge, then they continue in a way so that the changes of the load (in absolute values) on each of the two paths is minimized. Also note that every canonical path performs a random walk on G, i.e., at every incident matching edge it switches to the opposite node with probability 1/2, otherwise it stays at the current node.

Observation 4.6. *Fix the load vector at the end of round t_1 and consider the canonical path $\mathcal{P}_v = (\mathcal{P}_v^{(t_1)} = v, \mathcal{P}_v^{(t_1+1)}, \ldots)$ of v from round t_1. Then for any round $t \geqslant t_1$ and any node $w \in V$ it holds that*

$$\mathbf{Pr}\left[\,\mathcal{P}_v^{(t)} = w\,\right] = \mathbf{M}_{v,w}^{[t_1+1,t]}.$$



(a) $x_{v_t}^{(t)} \geqslant x_u^{(t)}$. (b) $x_{v_t}^{(t)} \geqslant x_u^{(t)}$.

(c) $x_{v_t}^{(t)} < x_u^{(t)}$. (d) $x_{v_t}^{(t)} < x_u^{(t)}$.

Figure 2: Illustration of the four cases of Definition 4.5 with $\{v_t, u\} \in \mathbf{M}^{(t+1)}$. The black arrows indicate the value of $\Phi_{u,v}^{(t+1)}$ (i.e., the direction of the excess token (if any)) and the grey directed edge is part of the canonical path.

Now we sketch the key ideas of the proof of Lemma 4.4. We use the polynomial potential function $\Phi^{(t)}$ that only considers nodes with load at least 11 (see (4.1)). Using the condition that the discrepancy of the load vector is at most $(\log n)^{\varepsilon_d}$, it follows directly that the initial value of the potential Φ is upper bounded by an almost linear function in n (Observation 4.7). To prove that Φ decreases, we consider phases of length $\beta := (\log n)^{\varepsilon_t}$, where $\varepsilon_t \in (0,1)$ is some small constant. In each such phase, we consider the canonical paths starting from nodes with load at least 11 together with the canonical paths starting from nodes with load at most 9.

To lower bound the probability that two canonical paths collide, we use the relation between canonical paths and random walks, i.e. as long as both canonical paths have not been connected by a matching edge, they evolve like independent random walks. Then, we sieve out those nodes with load at least 11 from which a canonical path has only a small probability to collide with a canonical path starting from a node with load at most 9 (Definition 4.9 and Lemma 4.10). Using this, we establish in Lemma 4.11 that there are indeed sufficiently many collisions, i.e., sufficiently many canonical paths with load at least 11 that collide with a canonical path with load at most 9 within the phase of length β. Clearly, not every collision between two canonical paths reduces the potential (since the load of one of the two canonical paths can change before), but as we argue in Lemma 4.11, the decrease of the potential is essentially the number of collisions divided by the length of the phase β.

The next observation provides an easy bound on the initial value of the potential function $\Phi^{(t)}$ defined by (4.1), exploiting the small discrepancy of the load vector.

Observation 4.7. *Consider any load vector $x^{(t)}$ with $\overline{x} \in [0,1)$ and discrepancy at most $(\log n)^{\varepsilon_d}$. Then $\Phi^{(t)} \leqslant n \cdot ((\log n)^{\varepsilon_d} + 1)^8$.*

Our goal is to consider phases of length β and prove that after each such phase, the expected value of the potential function drops. First, we define different conditions for a pair of node u and round t, which are necessary to cope with the inherent randomness of the balancing circuit and especially the random matching model.

Definition 4.8. *Let $\varepsilon_t > 0$ be any constant. For any node $u \in V$ and round $t \in \mathbb{N}$, define the following three conditions:*

- $\textsc{Cond1}(u,t):\quad x_u^{(t)} \geqslant 11.$
- $\textsc{Cond2}(u,t):\quad \left\|\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}\right\|_2^2 \leqslant (\log n)^{-\varepsilon_t/7}.$
- $\textsc{Cond3}(u,t):\quad \sum_{w \in V} \left(\mathbf{M}_{u,w}^{[t+1,t+\beta]} \sum_{v \in V} x_v^{(t)} \mathbf{M}_{w,v}^{[t+1,t+\beta]}\right) \geqslant 4.$

Moreover let $\mathcal{S}_1^{(t)}$ be the set of nodes u satisfying $\textsc{Cond}1(u,t)$. *The sets $\mathcal{S}_2^{(t)}$ and $\mathcal{S}_3^{(t)}$ are defined in the same way.*

Note that $\textsc{Cond}2(u,t)$ ensures that the local neighborhood around the node u with respect to the (directed) graph induced by the matchings within the time-interval $[t+1,t+\beta]$ expands sufficiently. With regards to $\textsc{Cond}3(u,t)$, recall that the probability distribution of the location of the canonical path of u in round $t+\beta$ is $\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}$. Moreover, for any $w \in V$ and fixed load vector $x^{(t)}$, $\sum_{v\in V} x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]}$ is the expected load on node w in round $t+\beta$. Hence, if $\textsc{Cond}3(u,t)$ holds, then at the (random) location of the canonical path in round $t+\beta$, the expected load is large, i.e., at least 4. We point out that although this argument gives the right intuition, it is not precisely true, since conditioning on the location of a canonical path affects the random orientations within the interval $[t+1,t+\beta]$, which in turn affects the load vector at the end of round $t+\beta$.

Definition 4.9. *A node $u \in V$ is called* bad *in round t, if node u satisfies* $\textsc{Cond}2(u,t)$ *and* $\textsc{Cond}3(u,t)$ *simultaneously. Let $\mathcal{B}^{(t)} \subseteq V$ be the set of all bad nodes in round t.*

In other words, a node u is bad if despite enjoying good expansion in the graph induced by the matchings within the time-interval $[t+1,t+\beta]$, the expected number of tokens at the endpoint of the canonical path is at least 4.

The next lemma shows that the number of bad nodes decreases exponentially in the phase length $\beta = (\log n)^{\varepsilon_t}$. Moreover the second statement provides a tail bound which is exponentially small in the "sparseness" σ.

Lemma 4.10. *Let $\varepsilon_t \in (0,1), \varepsilon_d \in (0,1)$ be two arbitrarily chosen constants. Fix an arbitrary load vector $x^{(0)}$ with discrepancy at most $(\log n)^{\varepsilon_d}$ and $\overline{x} \in [0,1)$. Then the following statements hold:*

- *For any round $t \geqslant \tau_{\mathrm{cont}}(n, n^{-3})$,*

$$\mathbf{Pr}\left[\left|\mathcal{B}^{(t)}\right| \leqslant n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_t/9}}\right] \geqslant 1 - \mathrm{e}^{-(\log n)^{\varepsilon_t/9}} - n^{-1}.$$

- *If $x^{(0)}$ is non-negative and satisfies $\|x^{(0)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{\sigma}}$ for some constant $\sigma \in (0,1)$, then for any round $t \geqslant \tau_{\mathrm{cont}}(n, n^{-3})$,*

$$\mathbf{Pr}\left[\left|\mathcal{B}^{(t)}\right| \leqslant n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_t/17+\sigma}}\right] \geqslant 1 - \mathrm{e}^{-(\log n)^{\varepsilon_t/17+\sigma}} - n^{-1}.$$

Proof. We begin by proving the first statement. Fix any node $u \in V$ in round t and consider $\left\|\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}\right\|_2^2$. While this is a deterministic value for the balancing circuit model, it is a random variable in the random matching model. By Lemma B.4,

$$\mathbf{Pr}\left[\left\|\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}\right\|_2^2 \geqslant (\log n)^{-\varepsilon_t/6}\right] \leqslant \mathrm{e}^{-(\log n)^{\varepsilon_t/2}}.$$

For the balancing circuit model, we replace the β rounds by $d \cdot \beta$ rounds which corresponds to the β-times iteration of the round matrix $\mathbf{M} = \prod_{i=1}^{d} \mathbf{M}^{(i)}$. Hence, by Corollary B.2 we have

$$\left\|\mathbf{M}_{u,\cdot}^{\beta}\right\|_2^2 \leqslant \left\|\mathbf{M}_{u,\cdot}^{\beta}\right\|_\infty \leqslant \frac{1}{n} + \mathcal{O}\left(\frac{1}{(\log n)^{\varepsilon_t/2}}\right) \leqslant (\log n)^{-\varepsilon_t/6}.$$

Let us now return to the random matching model. By Markov's inequality we get

$$\mathbf{Pr}\left[\left|\{u \in V \colon u \text{ does not satisfy } \textsc{Cond}2(u,t)\}\right| \geqslant n \cdot \mathrm{e}^{-\frac{1}{2}(\log n)^{\varepsilon_t/2}}\right] \leqslant \mathrm{e}^{-\frac{1}{2}\cdot(\log n)^{\varepsilon_t/2}},$$

that is,

$$\mathbf{Pr}\left[\left|V\setminus\mathcal{S}_2^{(t)}\right|\leqslant n\cdot \mathrm{e}^{-\frac{1}{2}\cdot(\log n)^{\varepsilon_t/2}}\right]\geqslant 1-\mathrm{e}^{-\frac{1}{2}(\log n)^{\varepsilon_t/2}}. \tag{4.2}$$

Note that $\mathcal{S}_2^{(t)}$ depends only on the random choices for the matchings within the time-interval $[t+1,t+\beta]$, which is independent of the load vector $x^{(t)}$ and the matchings in the time-interval $[0,t]$. Since $t\geqslant\tau_{\text{cont}}(n,n^{-3})$, it follows that

$$\mathbf{Pr}\left[\ [0,t]\text{ is }(n,n^{-3})\text{–smoothing }\right]\geqslant 1-n^{-1}. \tag{4.3}$$

For the remainder of the proof, we tacitly assume that $[0,t]$ is (n,n^{-3})–smoothing. Hence by Lemma 2.13, it holds for any node $u\in\mathcal{S}_2^{(t)}$ and $\delta>1/n$ that

$$\mathbf{Pr}\left[\left|\sum_{v\in V}x_v^{(t)}\mathbf{M}_{u,v}^{[t+1,t+\beta]}-\overline{x}\right|\geqslant\delta\ \middle|\ u\in\mathcal{S}_2^{(t)}\right]\leqslant 2\exp\left(-\frac{(\delta-1/(2n))^2}{4\cdot\left\|\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}\right\|_2^2}\right).$$

Choosing $\delta=1$ and recalling $\overline{x}\leqslant 1$ yields

$$\mathbf{Pr}\left[\sum_{v\in V}x_v^{(t)}\mathbf{M}_{u,v}^{[t+1,t+\beta]}\geqslant 2\ \middle|\ u\in\mathcal{S}_2^{(t)}\right]\leqslant 2\exp\left(-\frac{(1-1/(2n))^2}{4\cdot(\log n)^{-\varepsilon_t/7}}\right)\leqslant \mathrm{e}^{-(\log n)^{\varepsilon_t/8}}.$$

Hence by Markov's inequality,

$$\mathbf{Pr}\left[\left|\left\{u\in\mathcal{S}_2^{(t)}\colon\ \sum_{v\in V}x_v^{(t)}\mathbf{M}_{u,v}^{[t+1,t+\beta]}\geqslant 2\right\}\right|\geqslant\left|\mathcal{S}_2^{(t)}\right|\cdot\mathrm{e}^{-\frac{1}{2}\cdot(\log n)^{\varepsilon_t/8}}\right]\leqslant\mathrm{e}^{-\frac{1}{2}\cdot(\log n)^{\varepsilon_t/8}}.$$

Let $\mathcal{R}^{(t)}:=\left\{u\in V\colon\ \sum_{v\in V}x_v^{(t)}\mathbf{M}_{u,v}^{[t+1,t+\beta]}\geqslant 2\right\}$. Then we can rewrite the inequality above as

$$\mathbf{Pr}\left[\left|\mathcal{R}^{(t)}\cap\mathcal{S}_2^{(t)}\right|\leqslant\left|\mathcal{S}_2^{(t)}\right|\cdot\mathrm{e}^{-\frac{1}{2}\cdot(\log n)^{\varepsilon_t/8}}\right]\geqslant 1-\mathrm{e}^{-\frac{1}{2}(\log n)^{\varepsilon_t/8}}. \tag{4.4}$$

Combining (4.2), (4.4) and $\mathcal{R}^{(t)}\subseteq\left(\mathcal{R}^{(t)}\cap\mathcal{S}_2^{(t)}\right)\cup\left(V\setminus\mathcal{S}_2^{(t)}\right)$ gives

$$\mathbf{Pr}\left[\left|\mathcal{R}^{(t)}\right|\leqslant n\cdot\mathrm{e}^{-\frac{1}{2}(\log n)^{\varepsilon_t/8}}+n\cdot\mathrm{e}^{-\frac{1}{2}\cdot(\log n)^{\varepsilon_t/2}}\right]\geqslant 1-2\mathrm{e}^{-\frac{1}{2}\cdot(\log n)^{\varepsilon_t/8}}. \tag{4.5}$$

Our next goal is to upper bound the size of $\mathcal{S}_3^{(t)}$ in terms of $|\mathcal{R}^{(t)}|$. By definition, we have

$$\sum_{u\in\mathcal{S}_3^{(t)}}\sum_{w\in V}\left(\mathbf{M}_{u,w}^{[t+1,t+\beta]}\sum_{v\in V}x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right)\geqslant 4\cdot\left|\mathcal{S}_3^{(t)}\right|. \tag{4.6}$$

On the other hand, since $x_{\max}^{(t)}\leqslant x_{\max}^{(0)}\leqslant(\log n)^{\varepsilon_d}+1\leqslant 2(\log n)^{\varepsilon_d}$, we have

$$\begin{aligned}
&\sum_{u\in\mathcal{S}_3^{(t)}}\sum_{w\in V}\left(\mathbf{M}_{u,w}^{[t+1,t+\beta]}\sum_{v\in V}x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right)\\
&\leqslant\sum_{u\in\mathcal{S}_3^{(t)}}\sum_{w\in\mathcal{R}^{(t)}}\left(\mathbf{M}_{u,w}^{[t+1,t+\beta]}\sum_{v\in V}x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right)+\sum_{u\in\mathcal{S}_3^{(t)}}\sum_{w\in V\setminus\mathcal{R}^{(t)}}\left(\mathbf{M}_{u,w}^{[t+1,t+\beta]}\sum_{v\in V}x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right)\\
&\leqslant 2(\log n)^{\varepsilon_d}\cdot\sum_{u\in\mathcal{S}_3^{(t)}}\sum_{w\in\mathcal{R}^{(t)}}\mathbf{M}_{u,w}^{[t+1,t+\beta]}\sum_{v\in V}\mathbf{M}_{w,v}^{[t+1,t+\beta]}+2\left|\mathcal{S}_3^{(t)}\right|\\
&\leqslant 2(\log n)^{\varepsilon_d}\cdot\sum_{w\in\mathcal{R}^{(t)}}\sum_{u\in V}\mathbf{M}_{u,w}^{[t+1,t+\beta]}+2\left|\mathcal{S}_3^{(t)}\right|\\
&\leqslant 2(\log n)^{\varepsilon_d}\cdot\left|\mathcal{R}^{(t)}\right|+2\left|\mathcal{S}_3^{(t)}\right|.
\end{aligned} \tag{4.7}$$

Combining (4.6) and (4.7) we have

$$\left|\mathcal{S}_3^{(t)}\right| \leqslant (\log n)^{\varepsilon_d} \cdot \left|\mathcal{R}^{(t)}\right|.$$

Using this and $\mathcal{B}^{(t)} \subseteq \mathcal{S}_3^{(t)}$, the first statement follows from (4.5).

Let us now turn to the second statement of the lemma. As before, we tacitly assume that the time-interval $[0,t]$ is (n, n^{-3})–smoothing, which holds with probability at least $1-n^{-1}$. Fix any node $u \in \mathcal{S}_2^{(t)}$. Note that

$$\sum_{w\in V}\left(\mathbf{M}_{u,w}^{[t+1,t+\beta]}\sum_{v\in V} x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right) = \sum_{v\in V}\left(\sum_{w\in V}\mathbf{M}_{u,w}^{[t+1,t+\beta]}\cdot \mathbf{M}_{w,v}^{[t+1,t+\beta]}\right)\cdot x_v^{(t)}.$$

Now we apply Lemma 3.5 with $y_v = \sum_{w\in V}\mathbf{M}_{u,w}^{[t+1,t+\beta]}\cdot \mathbf{M}_{w,v}^{[t+1,t+\beta]}$. Since $u \in \mathcal{S}_2^{(t)}$, the vector y satisfies

$$\|y\|_\infty \leqslant \left\|\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}\right\|_\infty \leqslant \left\|\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}\right\|_2 \leqslant (\log n)^{-\varepsilon_t/14},$$

and choosing $\delta = \varepsilon_t/16$ in Lemma 3.5 gives

$$\begin{aligned}
&\mathbf{Pr}\left[\, u\in\mathcal{B}^{(t)}\,\right]\\
&= \mathbf{Pr}\left[\sum_{v\in V}\left(\sum_{w\in V}\mathbf{M}_{u,w}^{[t+1,t+\beta]}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right)\cdot x_v^{(t)} \geqslant 4 \,\middle|\, u\in\mathcal{S}_2^{(t)}\right]\cdot \mathbf{Pr}\left[\, u\in\mathcal{S}_2^{(t)}\,\right]\\
&\leqslant \mathbf{Pr}\left[\sum_{v\in V}\left(\sum_{w\in V}\mathbf{M}_{u,w}^{[t+1,t+\beta]}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right)\cdot x_v^{(t)} \geqslant \mathrm{e}^{-\frac{1}{5}(\log n)^{\sigma}} + 8(\log n)^{-\varepsilon_t/14}\cdot(\log n)^{\varepsilon_t/16} \,\middle|\, u\in\mathcal{S}_2^{(t)}\right]\\
&\leqslant \exp\left(-(\log n)^{\varepsilon_t/16+\sigma}/6\right) \leqslant \exp\left(-2\cdot(\log n)^{\varepsilon_t/17+\sigma}\right).
\end{aligned}$$

Therefore, by Markov's inequality,

$$\mathbf{Pr}\left[\,\left|\mathcal{B}^{(t)}\right| \geqslant n\cdot \mathrm{e}^{-(\log n)^{\varepsilon_t/17+\sigma}}\,\right] \leqslant \mathrm{e}^{-(\log n)^{\varepsilon_t/17+\sigma}}.$$

■

For any round t, define $\mathcal{S}_{12}^{(t)}$ to be the set of nodes u satisfying $\textsc{Cond}1(u,t)$ and $\textsc{Cond}2(u,t)$ and let

$$V_1^{(t)} := \mathcal{S}_{12}^{(t)} \setminus \mathcal{B}^{(t)},$$

be the set of nodes u satisfying $\textsc{Cond}1(u,t)$ and $\textsc{Cond}2(u,t)$, but not $\textsc{Cond}3(u,t)$. Intuitively, every node in $V_1^{(t)}$ has (i) large load, (ii) expanding neighborhood and (iii) small expected load at the location of the canonical path of u in round $t+\beta$. Hence if there are many nodes in $V_1^{(t)}$, then we would hope for a decrease in the potential Φ, which is formalized in the following lemma.

Lemma 4.11. *Let $\varepsilon_b, \varepsilon_d$ and ε_t be three arbitrary constants in the interval $(0,1)$. Fix a load vector $x^{(t)}$ with $\overline{x}\in[0,1)$ and discrepancy at most $(\log n)^{\varepsilon_d}$. Assume that*

$$\left|V_1^{(t)}\right| \geqslant \frac{1}{2}\left|\mathcal{S}_1^{(t)}\right| - n\cdot \mathrm{e}^{-(\log n)^{\varepsilon_b}}$$

Then it holds in round $t+\beta, \beta := (\log n)^{\varepsilon_t}$, that

$$\mathbf{E}\left[\,\Phi^{(t+\beta)}\,\right] \leqslant \max\left\{\left(1-\frac{1}{18(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)\cdot\Phi^{(t)},\, 4n\cdot \mathrm{e}^{-(\log n)^{\varepsilon_b}}\cdot\left((\log n)^{\varepsilon_d}+1\right)^8\right\}.$$

Proof. We remark that the precondition on $V_1^{(t)}$ depends only on the random choices for the matchings in the interval $[t+1,t+\beta]$, but not on the orientation of these edges. Therefore, the orientations of the matchings in the time-interval $[t+1,t+\beta]$ are still chosen independently and uniformly at random.

By definition, the node $u \in V_1^{(t)}$ satisfies

$$\sum_{w\in W}\left(\mathbf{M}_{u,w}^{[t+1,t+\beta]}\sum_{v\in V} x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]}\right) < 4.$$

Let

$$\mathcal{G}^{(t)} := \left\{ w \in V\colon \sum_{v\in V} x_v^{(t)}\mathbf{M}_{w,v}^{[t+1,t+\beta]} \leqslant 8 \right\}.$$

Since the probability distribution of $\mathcal{P}_u^{(t+\beta)}$, the location of the canonical path of u from round t in round $t+\beta$, is given by $\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}$, it follows by Markov's inequality that

$$\mathbf{Pr}\left[\,\mathcal{P}_u^{(t+\beta)} \in \mathcal{G}^{(t)}\,\right] \geqslant \frac{1}{2}. \tag{4.8}$$

Fix now a node $w \in \mathcal{G}^{(t)}$. Our next aim is to derive a lower bound on the probability that there is a canonical path starting from a node k with $x_k^{(t)} \leqslant 9$ that reaches w in round $t+\beta$. To this end, let us define $V_2^{(t)} := \left\{ v \in V\colon x_v^{(t)} \leqslant 9 \right\}$ and $\alpha_w := \sum_{v\in V_2^{(t)}} \mathbf{M}_{v,w}^{[t+1,t+\beta]} = \sum_{v\in V_2^{(t)}} \mathbf{M}_{w,v}^{[t+1,t+\beta]}$ for node w. Then,

$$\begin{aligned}
8 &\geqslant \sum_{v\in V} x_v^{(t)} \cdot \mathbf{M}_{v,w}^{[t+1,t+\beta]}\\
&= \sum_{v\in V_2^{(t)}} x_v^{(t)} \cdot \mathbf{M}_{v,w}^{[t+1,t+\beta]} + \sum_{v\in V\setminus V_2^{(t)}} x_v^{(t)} \cdot \mathbf{M}_{v,w}^{[t+1,t+\beta]}\\
&\geqslant -(\log n)^{\varepsilon_d}\cdot\alpha_w + 10\cdot(1-\alpha_w),
\end{aligned}$$

and rearranging yields for any $w \in \mathcal{G}^{(t)}$,

$$\alpha_w \geqslant \frac{2}{(\log n)^{\varepsilon_d}+10} \geqslant \frac{1}{(\log n)^{\varepsilon_d}}. \tag{4.9}$$

Let us now lower bound the probability for the event that the canonical path starting from $u \in V_1^{(t)}$ collides with a canonical path starting from a node in $V_2^{(t)}$ within the time-interval $[t+1,t+\beta]$. By (4.8), the canonical path of u reaches a node $w \in \mathcal{G}^{(t)}$ with probability at least $1/2$. Moreover, a canonical path starting from a fixed node $v \in V_2^{(t)}$ performs an independent random walk based on the random orientations as long as it does not reach a node g in round $s \in [t+1,t+\beta]$ so that node g is matched with another node g' in round s and the canonical path of u is located at node g' in round $s-1$. Hence,

$$\mathbf{Pr}\left[\exists s \in [t+1,t+\beta]\colon \left\{\mathcal{P}_u^{(s)},\mathcal{P}_v^{(s)}\right\} \in \mathbf{M}^{(s)} \;\middle|\; \mathcal{P}_u^{(t+\beta)} = w\right] \geqslant \mathbf{M}_{v,w}^{[t+1,t+\beta]}.$$

i.e., the probability that the canonical path starting from a fixed node $v \in V_2^{(t)}$ collides with the canonical path starting from u in round t and reaches node w in round $t+\beta$, is at least $\mathbf{M}_{v,w}^{[t+1,t+\beta]}$. See Figure 3 for an illustration.

We now define a bipartite graph $H = \left(V_1^{(t)} \cup V_2^{(t)}, E\right)$. We place an edge $\{u,v\} \in E$ if the canonical path of $u \in V_1^{(t)}$ collides with the canonical path of $v \in V_2^{(t)}$ within the time-interval

$[t+1, t+\beta]$. We first lower bound the expected number of edges in H:

$$
\begin{aligned}
&\mathbf{E}\left[\,|E(H)|\,\right] \\
&= \sum_{u \in V_1^{(t)}} \sum_{v \in V_2^{(t)}} \mathbf{Pr}\left[\exists s \in [t+1, t+\beta] \colon \left\{\mathcal{P}_u^{(s)}, \mathcal{P}_v^{(s)}\right\} \in \mathbf{M}^{(s)}\right] \\
&= \sum_{u \in V_1^{(t_1)}} \sum_{v \in V_2^{(t)}} \sum_{w \in V} \mathbf{Pr}\left[\mathcal{P}_u^{(t+\beta)} = w\right] \cdot \mathbf{Pr}\left[\exists s \in [t+1, t+\beta] \colon \left\{\mathcal{P}_u^{(s)}, \mathcal{P}_v^{(s)}\right\} \in \mathbf{M}^{(s)} \;\middle|\; \mathcal{P}_u^{(t+\beta)} = w\right] \\
&\geqslant \sum_{u \in V_1^{(t)}} \sum_{v \in V_2^{(t)}} \sum_{w \in \mathcal{G}^{(t)}} \mathbf{Pr}\left[\mathcal{P}_u^{(t+\beta)} = w\right] \cdot \mathbf{M}_{v,w}^{[t+1,t+\beta]} \\
&= \sum_{u \in V_1^{(t)}} \sum_{w \in \mathcal{G}^{(t)}} \mathbf{Pr}\left[\mathcal{P}_u^{(t+\beta)} = w\right] \cdot \sum_{v \in V_2^{(t)}} \mathbf{M}_{v,w}^{[t+1,t+\beta]}.
\end{aligned}
$$

By (4.9), we know that $\sum_{v \in V_2^{(t)}} \mathbf{M}_{v,w}^{[t+1,t+\beta]} \geqslant \frac{1}{(\log n)^{\varepsilon_d}}$. Hence

$$
\mathbf{E}\left[\,|E(H)|\,\right] \geqslant \frac{1}{(\log n)^{\varepsilon_d}} \cdot \sum_{u \in V_1^{(t)}} \sum_{w \in \mathcal{G}^{(t)}} \mathbf{Pr}\left[\mathcal{P}_u^{(t+\beta)} = w\right] \geqslant \frac{1}{(\log n)^{\varepsilon_d}} \cdot \left|V_1^{(t)}\right| \cdot \frac{1}{2}\,,
$$

where the second inequality follows from (4.8). By the precondition of this lemma,

$$
\left|V_1^{(t)}\right| \geqslant \frac{1}{2} \cdot \left|\mathcal{S}_1^{(t)}\right| - n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_b}}.
$$

If $\left|\mathcal{S}_1^{(t)}\right| \geqslant 4n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_b}}$, then by the above inequality we have $\left|V_1^{(t)}\right| \geqslant \frac{1}{4}\left|\mathcal{S}_1^{(t)}\right|$. Otherwise,

$$
\Phi^{(t)} = \sum_{\substack{u \in V \\ x_u^{(t)} \geqslant 11}} \left(x_u^{(t)}\right)^8 \leqslant \left|\mathcal{S}_1^{(t)}\right| \cdot \left((\log n)^{\varepsilon_d} + 1\right)^8 \leqslant 4n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_b}} \cdot \left((\log n)^{\varepsilon_d} + 1\right)^8,
$$

which finishes the proof of this lemma, as $\Phi^{(t+\beta)} \leqslant \Phi^{(t)}$. Hence it remains to consider the case $\left|\mathcal{S}_1^{(t)}\right| \geqslant 4n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_b}}$ (and consequently, $\left|V_1^{(t)}\right| \geqslant \frac{1}{4}\left|\mathcal{S}_1^{(t)}\right|$). This allows us to lower bound $\mathbf{E}\left[\,|E(H)|\,\right]$ as follows:

$$
\begin{aligned}
\mathbf{E}\left[\,|E(H)|\,\right] &\geqslant \frac{1}{8 \cdot (\log n)^{\varepsilon_d}} \cdot \left|\mathcal{S}_1^{(t)}\right| \\
&\geqslant \frac{1}{8 \cdot (\log n)^{\varepsilon_d}} \cdot \Phi^{(t)} \cdot \frac{1}{\left((\log n)^{\varepsilon_d} + 1\right)^8} \\
&\geqslant \frac{1}{9 \cdot (\log n)^{9\varepsilon_d}} \cdot \Phi^{(t)},
\end{aligned} \tag{4.10}
$$

for sufficiently large n.

Consider now the auxiliary graph H again. By definition, H contains all those edges $\{u, v\}$ with the property that two canonical paths of $u \in V_1^{(t)}$ and $v \in V_2^{(t)}$ collide within the time-interval $[t+1, t+\beta]$. Fix now an arbitrary edge $\{u, v\} \in E(H)$ and let $s \in [t+1, t+\beta]$ be the round when the canonical paths of u and v meet for the first time. Then in round s these two canonical paths contribute a decrease of at least one to the potential function Φ if $x_u^{(s)} \geqslant 11$ and $x_v^{(s)} \leqslant 9$. If this does not happen, then we know that either $x_u^{(s)} < 11$ or $x_v^{(s)} > 9$ which means that one of the two canonical paths has contributed a decrease of at least one before round s. Put differently, every node in $V_1^{(t)}$ which is not isolated in H contributes at least $1/2$ to the reduction of the potential function. Since the maximum degree of graph H, denoted by $\Delta(H)$,

is at most $\beta = (\log n)^{\varepsilon_t}$, it follows that the number of nodes in $V_1^{(t)}$ which are not isolated is at least $|E(H)|/(\log n)^{\varepsilon_t}$. Therefore,

$$\Phi^{(t)} - \Phi^{(t+\beta)} \geqslant \frac{1}{2} \cdot \frac{|E(H)|}{\Delta(H)} \geqslant \frac{|E(H)|}{2 \cdot (\log n)^{\varepsilon_t}}.$$

Taking expectations on both sides and using our lower bound from (4.10) on $\mathbf{E}\left[\, |E(H)| \,\right]$ finally gives

$$\mathbf{E}\left[\Phi^{(t)} - \Phi^{(t+\beta)} \right] \geqslant \frac{\mathbf{E}\left[\, |E(H)| \,\right]}{2 \cdot (\log n)^{\varepsilon_t}} \geqslant \frac{1}{18 \cdot (\log n)^{\varepsilon_t + 9\varepsilon_d}} \cdot \Phi^{(t)},$$

which completes the proof. ■

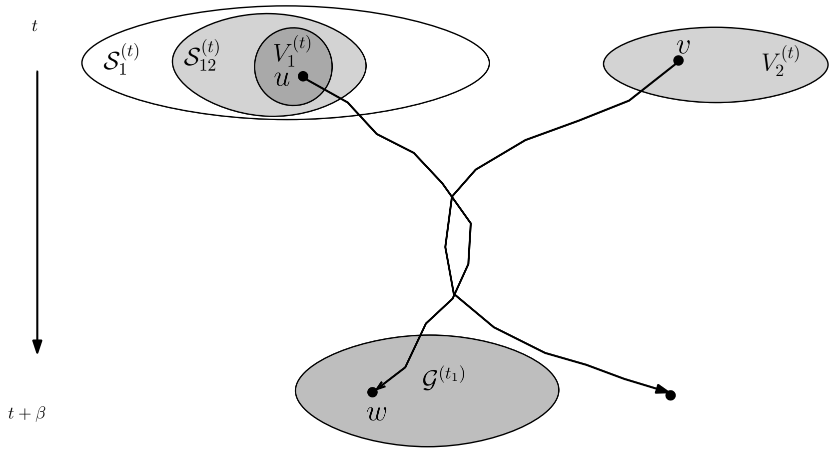


Figure 3: An illustration of a collision between the canonical path starting from $u \in V_1^{(t)}$ in round t and a canonical path of $v \in V_2^{(t)}$ from round t, where the canonical path of u reaches a node $w \in \mathcal{G}^{(t)}$ in round $t + \beta$.

Finally, we are now ready to prove Lemma 4.4 by combining Observation 4.7, Lemma 4.10 and Lemma 4.11.

Proof of Lemma 4.4. We begin with the first statement. Let $t_1 := t + \tau_{\text{cont}}(n, n^{-3}) = t + \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$. By Observation 4.7 it holds that

$$\Phi^{(t_1)} \leqslant \Phi^{(t)} \leqslant n \cdot ((\log n)^{\varepsilon_d} + 1)^8 .$$

For any round $s \in [t_1, t_1 + \log n]$, let $\mathcal{E}_s$ be the event defined by

$$\mathcal{E}_s := \left\{ \left| V_1^{(s)} \right| \geqslant \frac{1}{2} \left| \mathcal{S}_1^{(s)} \right| - n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_b}} \right\},$$

where $\varepsilon_b := \varepsilon_t/9$. Further, let $\mathcal{E} := \bigwedge_{s=t_1}^{t_1+\log n} \mathcal{E}_s$. We are interested in a lower bound on $\mathbf{Pr}\left[\, \mathcal{E} \,\right]$. First, by the first statement of Lemma 4.10, for any round $s \geqslant t_1$:

$$\mathbf{Pr}\left[\left| \mathcal{B}^{(s)} \right| \leqslant n \cdot \mathrm{e}^{-(\log n)^{\varepsilon_t/9}} \right] \geqslant 1 - \mathrm{e}^{-(\log n)^{\varepsilon_t/9}} - n^{-1}.$$

Moreover, for any $s \geqslant t_1$, Lemma B.4 gives for any $u \in \mathcal{S}_1^{(s)}$,

$$\mathbf{Pr}\left[\left\| \mathbf{M}_{u,\cdot}^{[s+1, s+(\log n)^{\varepsilon_t}]} \right\|_2^2 \geqslant (\log n)^{-\varepsilon_t/7} \right] \leqslant \mathrm{e}^{-(\log n)^{\varepsilon_t/2}},$$

Hence $\mathbf{E}\left[\left|\mathcal{S}_{12}^{(s)}\right|\right] \geqslant \left|\mathcal{S}_{1}^{(s)}\right| \cdot \left(1-\mathrm{e}^{-(\log n)^{\varepsilon_t/2}}\right)$ and $\mathbf{E}\left[\left|\mathcal{S}_{1}^{(s)} \setminus \mathcal{S}_{12}^{(s)}\right|\right] \leqslant \left|\mathcal{S}_{1}^{(s)}\right| \cdot \mathrm{e}^{-(\log n)^{\varepsilon_t/2}}$. Then by Markov's inequality,

$$\mathbf{Pr}\left[\left|\mathcal{S}_{12}^{(s)}\right| \leqslant \frac{1}{2}\cdot\left|\mathcal{S}_{1}^{(s)}\right|\right] \leqslant \mathbf{Pr}\left[\left|\mathcal{S}_{1}^{(s)} \setminus \mathcal{S}_{12}^{(s)}\right| \geqslant \frac{1}{2}\cdot \mathrm{e}^{(\log n)^{\varepsilon_t/2}} \cdot \mathbf{E}\left[\left|\mathcal{S}_{1}^{(s)} \setminus \mathcal{S}_{12}^{(s)}\right|\right]\right] \leqslant 2\mathrm{e}^{-(\log n)^{\varepsilon_t/2}},$$

and consequently

$$\mathbf{Pr}\left[\left|\mathcal{S}_{12}^{(s)}\right| \geqslant \frac{1}{2}\cdot\left|\mathcal{S}_{1}^{(s)}\right|\right] \geqslant 1 - 2\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/2}}. \tag{4.11}$$

Since $V_1^{(s)} := \mathcal{S}_{12}^{(s)} \setminus \mathcal{B}^{(s)}$, we conclude by the union bound that with probability at least $1 - \mathrm{e}^{-(\log n)^{\varepsilon_t/9}} - n^{-1} - 2\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/2}} \geqslant 1 - 2\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/9}}$,

$$\left|V_1^{(s)}\right| \geqslant \left|\mathcal{S}_{12}^{(s)}\right| - \left|\mathcal{B}^{(s)}\right| \geqslant \frac{1}{2}\cdot\left|\mathcal{S}_{1}^{(s)}\right| - n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/9}}.$$

Hence by the union bound over $\log n$ rounds,

$$\mathbf{Pr}\left[\,\mathcal{E}\,\right] \geqslant 1 - \log n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/9}} \geqslant 1 - \mathrm{e}^{-(\log n)^{\varepsilon_t/10}}. \tag{4.12}$$

By Lemma 4.11 with $\varepsilon_b = \varepsilon_t/9$, if $\Phi^{(t_1)} \geqslant n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/10}}$, then

$$\left(1 - \frac{1}{18\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)\cdot\Phi^{(t_1)} \geqslant 4n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_b}}\cdot\left((\log n)^{\varepsilon_d}+1\right)^8$$

and

$$\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}_{t_1}\right] \leqslant \left(1 - \frac{1}{18\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)\cdot\Phi^{(t_1)}, \tag{4.13}$$

Our claim is that a very similar inequality also holds for $\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \mid \mathcal{E}\,\right]$, whenever $\Phi^{(t_1)} \geqslant n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/10}}$. Note that

$$\begin{aligned}
\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}_{t_1}\right] &\geqslant \mathbf{Pr}\left[\,\mathcal{E} \mid \mathcal{E}_{t_1}\,\right]\cdot\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}\wedge\mathcal{E}_{t_1}\right] \\
&\geqslant \mathbf{Pr}\left[\,\mathcal{E}\wedge\mathcal{E}_{t_1}\,\right]\cdot\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}\right] \\
&= \mathbf{Pr}\left[\,\mathcal{E}\,\right]\cdot\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}\right] \\
&\geqslant \left(1-\mathrm{e}^{-(\log n)^{\varepsilon_t/10}}\right)\cdot\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}\right],
\end{aligned}$$

where the last inequality follows from (4.12). In combination with (4.13), we have

$$\begin{aligned}
\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}\right] &\leqslant \left(1-\mathrm{e}^{-(\log n)^{\varepsilon_t/10}}\right)^{-1}\cdot\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}_{t_1}\right] \\
&\leqslant \left(1-\mathrm{e}^{-(\log n)^{\varepsilon_t/10}}\right)^{-1}\cdot\left(1 - \frac{1}{18\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)\cdot\Phi^{(t_1)} \\
&\leqslant \left(1 - \frac{1}{36\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)\cdot\Phi^{(t_1)}
\end{aligned} \tag{4.14}$$

as long as $\Phi^{(t_1)} \geqslant n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/10}}$.

Because of (4.12), we will work in the following on the probability space conditioned on $\mathcal{E} = \bigwedge_{s=t_1}^{t_1+\log n}\mathcal{E}_s$. Our next aim is to iterate (4.14). For simplicity, we shall iterate the following inequality:

$$\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}\right] \leqslant \left(1 - \frac{1}{36\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)\cdot\Phi^{(t_1)} + n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/10}},$$

which holds for arbitrary $\Phi^{(t_1)}$. Applying the chain rule of expectations $(\gamma-1)$ times gives

$$\mathbf{E}\left[\,\Phi^{(t_1+\gamma\cdot\beta)} \;\middle|\; \mathcal{E}\,\right] \leqslant \left(1-\frac{1}{36\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)^{\gamma}\cdot\Phi^{(t_1)}+\gamma\cdot n\cdot \mathrm{e}^{-(\log n)^{\varepsilon_t/10}}.$$

Since the initial potential satisfies $\Phi^{(t_1)}\leqslant n\cdot((\log n)^{\varepsilon_d}+1)^8$ (cf. Observation 4.7), choosing $\gamma=(\log n)^{1-\varepsilon_t}$ yields

$$\begin{aligned}\mathbf{E}\left[\,\Phi^{(t_1+\gamma\cdot\beta)} \;\middle|\; \mathcal{E}\,\right] &\leqslant \left(1-\frac{1}{36\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)^{\gamma}\cdot\Phi^{(t_1)}+\gamma\cdot n\cdot \mathrm{e}^{-(\log n)^{\varepsilon_t/10}}\\ &\leqslant \exp\left(-\frac{1}{36}\cdot(\log n)^{1-2\varepsilon_t-9\varepsilon_d}\right)\cdot n\cdot((\log n)^{\varepsilon_d}+1)^8+\frac{1}{2}\cdot n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/11}}\\ &\leqslant \frac{1}{2}\cdot\exp\left(-(\log n)^{1-2\varepsilon_t-10\varepsilon_d}\right)\cdot n+\frac{1}{2}\cdot n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/11}}.\end{aligned}$$

Solving $1-2\varepsilon_t-10\varepsilon_d=\varepsilon_t/11$ yields $\varepsilon_t=\frac{1-10\varepsilon_d}{23/11}$, for which the expected value of the potential in round $t_1+\gamma\cdot(\log n)^{\varepsilon_t}=t_1+\log n$ is at most

$$\mathbf{E}\left[\,\Phi^{(t_1+\log n)} \;\middle|\; \mathcal{E}\,\right]\leqslant n\cdot\exp\left(-(\log n)^{\frac{1-10\varepsilon_d}{23}}\right).$$

Hence by choosing ε_d small enough and by Markov's inequality,

$$\mathbf{Pr}\left[\,\Phi^{(t_1+\log n)}\geqslant n\cdot\mathrm{e}^{-(\log n)^{\frac{1}{24}}} \;\middle|\; \mathcal{E}\,\right]\leqslant \mathrm{e}^{-(\log n)^{\Omega(1)}}. \tag{4.15}$$

Since $t_1+\log n=t+\tau_{\text{cont}}(n,n^{-3})+\log n=t+\mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$, taking the union bound over (4.12) and (4.15) yields the first statement of Lemma 4.4.

Now we prove the second statement of Lemma 4.4. Since the load vector $x^{(t_1)}$ satisfies $\|x^{(t_1)}\|_1\leqslant n\cdot\mathrm{e}^{-(\log n)^{\sigma}}$, by the second statement of Lemma 4.10 it holds for any round $s\geqslant t_1$ that

$$\mathbf{Pr}\left[\,\left|\mathcal{B}^{(s)}\right|\leqslant n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/17+\sigma}}\,\right]\geqslant 1-\mathrm{e}^{-(\log n)^{\varepsilon_t/17+\sigma}}-\frac{1}{n}. \tag{4.16}$$

For any round $s\in[t_1,t_1+\log n]$, redefine the event $\mathcal{E}_s$ as

$$\mathcal{E}_s:=\left\{\left|V_1^{(s)}\right|\geqslant\frac{1}{2}\left|\mathcal{S}_1^{(s)}\right|-n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_b}}\right\},$$

where $\varepsilon_b:=\varepsilon_t/17+\sigma$, and let $\mathcal{E}:=\bigwedge_{s=t_1}^{t_1+\log n}\mathcal{E}_s$. Combining (4.11) with (4.16),

$$\mathbf{Pr}\left[\,\mathcal{E}\,\right]\geqslant 1-\log n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/17+\sigma}}-2\log n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/9}}-\log n\cdot\frac{1}{n}=1-\mathrm{e}^{-(\log n)^{\Omega(1)}}.$$

Then applying Lemma 4.11 with $\varepsilon_b=\varepsilon_t/17+\sigma$ and (4.14) above gives

$$\mathbf{E}\left[\,\Phi^{(t_1+\beta)} \;\middle|\; \mathcal{E}\,\right]\leqslant\left(1-\frac{1}{36\cdot(\log n)^{\varepsilon_t+9\varepsilon_d}}\right)\cdot\Phi^{(t_1)}, \tag{4.17}$$

as long as $\Phi^{(t_1)}\geqslant n\cdot\mathrm{e}^{-(\log n)^{\varepsilon_t/18+\sigma}}$. As we did from (4.14), we conclude from (4.17) with $\gamma:=(\log n)^{1-\varepsilon_t}$ that

$$\begin{aligned}&\mathbf{E}\left[\,\Phi^{(t_1+\gamma\cdot\beta)} \;\middle|\; \mathcal{E}\,\right]\\ &\leqslant\exp\left(-\frac{1}{36}\cdot(\log n)^{1-2\varepsilon_t-9\varepsilon_d}\right)\cdot n\cdot((\log n)^{\varepsilon_d}+1)^8+\gamma\cdot n\cdot\exp\left(-(\log n)^{\varepsilon_t/18+\sigma}\right)\\ &\leqslant\frac{1}{2}\cdot n\cdot\exp\left(-(\log n)^{1-2\varepsilon_t-10\varepsilon_d}\right)+\frac{1}{2}\cdot n\cdot\exp\left(-(\log n)^{\varepsilon_t/19+\sigma}\right)\end{aligned}$$

Solving the equation

$$1 - 2\varepsilon_t - 10\varepsilon_d = \varepsilon_t/19 + \sigma$$

gives $\varepsilon_t = (1 - 10\varepsilon_d - \sigma) \cdot \frac{19}{39}$ for which

$$\begin{aligned}\mathbf{E}\left[\Phi^{(t_1+\gamma\cdot\beta)} \mid \mathcal{E}\right] &\leqslant n \cdot \exp\left(-(\log n)^{1-(1-10\varepsilon_d-\sigma)\cdot\frac{38}{39}-10\varepsilon_d}\right) \\ &= n \cdot \exp\left(-(\log n)^{1-10\varepsilon_d-\frac{38}{39}\cdot(1-10\varepsilon_d-\sigma)}\right).\end{aligned}$$

Using Markov's inequality,

$$\mathbf{Pr}\left[\Phi^{(t_1+\log n)} \geqslant n \cdot \exp\left(-(\log n)^{1-11\varepsilon_d-\frac{38}{39}\cdot(1-11\varepsilon_d-\sigma)}\right) \mid \mathcal{E}\right] \leqslant \mathrm{e}^{-(\log n)^{\Omega(1)}}.$$

Since $t_1 + \log n = t + \tau_{\text{cont}}(n, n^{-3}) + \log n = t + \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$, taking the union bound over (4.12) and the inequality above yields the second statement. ■

4.3 Proof of Theorem 4.3

We start with an outline of the proof of Theorem 4.3 (see also the illustration in Figure 4). Some techniques used here are similar to the ones used in proving Theorem 4.2. For instance, we also use a potential function and short phases of (poly-)logarithmic length to show that the potential decreases. However, to prove Theorem 4.3, we have to eliminate *all* tokens above a constant threshold, while in Theorem 4.2 it is enough to make this number smaller than $n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$, which is still polynomial in n. To achieve this improvement, we switch to a token-based viewpoint and exploit the sparseness of the load vector by using the approach from Section 3.

First, fix a node u with $x_u^{(t)} \geqslant 2$ and consider the next β rounds. For the proof outline, we confine ourselves to the case where the degree of G is small (the case with larger degree is actually slightly simpler). Clearly, the number of nodes from which a token could meet with a token in $B_\beta(u)$, i.e., the set of nodes with distance at most β to u, is at most $d^{2\beta}$. Now if β is small enough, then a relatively straightforward calculation shows that the total number of tokens on nodes in $B_\beta(u)$ is at most $16(\log n)^\varepsilon$ (Lemma 4.13). Moreover, if the nodes that host these $16(\log n)^\varepsilon$ tokens expand in the graph induced by the random matchings, then we obtain a good upper bound on the probability that *any* pair of these $16(\log n)^\varepsilon$ tokens meet in round $t+\beta$. This implies that the load along the canonical path starting from u in round t will decrease within the next β rounds.

Formally, we use an exponential potential function $\Lambda^{(t)}$ defined by

$$\Lambda^{(t)} := \sum_{u \in V} \Lambda_u^{(t)},$$

where

$$\Lambda_u^{(t)} := \begin{cases} \mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}\cdot x_u^{(t)}} & \text{if } x_u^{(t)} \geqslant 2 \\ 0 & \text{otherwise,} \end{cases}$$

and $\varepsilon > 0$ is a sufficiently small constant that will be fixed later. We shall prove that Λ drops by a factor of $\mathrm{e}^{\Omega(\beta\cdot(1-\lambda))}$ within β rounds. Since initially, Λ is polynomial in n with high probability (Lemma 4.12), it follows that after $\mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$ rounds, the value of Λ becomes zero with high probability and hence the maximum load is at most one.

We first derive some basic properties of this potential function. After that, we turn to the more involved task of establishing an expected drop of the potential.

Lemma 4.12. *Let $x^{(0)}$ be an arbitrary, non-negative load vector. Then the following holds:*

- *If $t \geqslant \tau_{\text{cont}}(n, n^{-2})$ and $\|x^{(0)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$, then with probability at least $1 - 2n^{-1}$, $\Lambda^{(t)} \leqslant 9n^2$.*
- *If two nodes u and v are matched in round t and $x_u^{(t-1)} - x_v^{(t-1)} \geqslant 2$, then*

$$\Lambda_u^{(t-1)} + \Lambda_v^{(t-1)} - \Lambda_u^{(t)} - \Lambda_v^{(t)} \geqslant \Lambda_u^{(t-1)} \cdot \left(1 - \frac{2}{\mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}}}\right).$$

- *For any $u \in V$, the function $\Lambda_u^{(t)}$ is convex in $x_u^{(t)}$, and hence $\Lambda^{(t)}$ is non-increasing in t.*

Proof. We start with the proof of the first statement. Since $t \geqslant \tau_{\text{cont}}(n, n^{-2})$, the time-interval $[0, t]$ is (n, n^{-2})–smoothing with probability at least $1 - n^{-1}$. Fix now all the matchings in $[0, t]$ and consider the orientations of the matching edges in the time-interval $[0, t]$, which, together with $x^{(0)}$, determine the load vector $x^{(t)}$. Fix any node $u \in V$ and consider $x_u^{(t)}$. By Lemma 3.1 and Lemma B.5,

$$\mathbf{E}\left[x_u^{(t)}\right] \leqslant 2 \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}} < 1.$$

By Lemma 3.3 we have for any $\delta > 9$,

$$\mathbf{Pr}\left[x_u^{(t)} \geqslant (1+\delta)\mathbf{E}\left[x_u^{(t)}\right]\right] \leqslant \delta^{-\delta\cdot\mathbf{E}\left[x_u^{(t)}\right]/2}.$$

Choosing $\delta = \rho/\mathbf{E}\left[x_u^{(t)}\right]$ for any real number $\rho \geqslant 1$ yields

$$\mathbf{Pr}\left[x_u^{(t)} \geqslant 2\rho\right] \leqslant \left(\mathbf{E}\left[x_u^{(t)}\right]\right)^{\rho/2} \leqslant 2^{\rho/2} \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}\cdot\rho/2}.$$

Hence,

$$\begin{aligned}
\mathbf{E}\left[\Lambda_u^{(t)}\right] &= \sum_{k=2}^{\infty} \mathbf{Pr}\left[x_u^{(t)} = k\right] \cdot \mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}\cdot k} \\
&\leqslant \sum_{k=2}^{\infty} \mathbf{Pr}\left[x_u^{(t)} \geqslant k\right] \cdot \mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}\cdot k} \\
&\leqslant \sum_{k=2}^{\infty} \mathrm{e}^{-(\log n)^{1-\varepsilon}\cdot k/4} \cdot \mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}\cdot k} \\
&\leqslant \sum_{k=2}^{\infty} \mathrm{e}^{-k/8} \leqslant \frac{1}{1 - \mathrm{e}^{-1/8}} < 9,
\end{aligned}$$

where the third inequality holds for sufficiently large n. Therefore, $\mathbf{E}\left[\Lambda^{(t)}\right] = \sum_{u \in V} \mathbf{E}\left[\Lambda_u^{(t)}\right] \leqslant 9n$ and by Markov's inequality,

$$\mathbf{Pr}\left[\Lambda^{(t)} \geqslant 9n^2\right] \leqslant n^{-1}.$$

This completes the first statement. For the second statement,

$$\begin{aligned}
\Lambda_u^{(t-1)} + \Lambda_v^{(t-1)} - \Lambda_u^{(t)} - \Lambda_v^{(t)} &\geqslant \Lambda_u^{(t-1)} \cdot \left(1 - \frac{\Lambda_u^{(t)} + \Lambda_v^{(t)} - \Lambda_v^{(t-1)}}{\Lambda_u^{(t-1)}}\right) \\
&\geqslant \Lambda_u^{(t-1)} \cdot \left(1 - \frac{\Lambda_u^{(t)} + \Lambda_v^{(t)}}{\Lambda_u^{(t-1)}}\right) \\
&\geqslant \Lambda_u^{(t-1)} \cdot \left(1 - \frac{2 \cdot \mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}\cdot(x_u^{(t-1)}-1)}}{\mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}\cdot(x_u^{(t-1)})}}\right) \\
&= \Lambda_u^{(t-1)} \cdot \left(1 - \frac{2}{\mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}}}\right),
\end{aligned}$$

where in the third inequality we used the fact that $x_u^{(t)} \leqslant x_u^{(t-1)} - 1$ and $x_v^{(t)} \leqslant x_u^{(t-1)} - 1$.

For the third statement, it suffices to prove that $\Lambda_u^{(t)}$ is convex which holds since $x \mapsto \mathrm{e}^{\frac{1}{8}(\log n)^{1-\varepsilon} \cdot x}$ is convex and $2 \cdot \mathrm{e}^{\frac{1}{8}(\log n)^{1-\varepsilon} \cdot 1} \leqslant \mathrm{e}^{\frac{1}{8}(\log n)^{1-\varepsilon} \cdot 2}$, which holds for sufficiently large n. ■

For the remainder of the proof, define $t_0 := \tau_{\mathrm{cont}}(n, n^{-2}) = \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$. Roughly speaking, these t_0 rounds are used to ensure a coarse balancing of the sparse initial load vector. More precisely, the next lemma proves that every small-sized set S contains only a polylogarithmic number of tokens under the condition that the initial load vector is sparse.

Lemma 4.13. *Consider the random matching model. Fix an arbitrary, non-negative load vector with $\|x^{(0)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$, where $\varepsilon > 0$ is a sufficiently small constant. Assume that the time-interval $[0, t]$ is (n, n^{-2})–smoothing. Then, for any subset of nodes $S \subseteq V$ with $|S| \leqslant 4 \cdot \mathrm{e}^{\frac{1}{2} \cdot (\log n)^{1-\varepsilon}}$ it holds that*

$$\mathbf{Pr}\left[\sum_{u \in S} x_u^{(t)} \geqslant 16 \cdot (\log n)^{\varepsilon}\right] \leqslant n^{-7}.$$

Proof. Let us consider the total number of tokens located in S in round t. Define

$$Z := \sum_{i=1}^{\|x^{(0)}\|_1} \mathbf{1}_{w_i^{(t)} \in S}.$$

Since the time-interval $[0, t]$ is (n, n^{-2})–smoothing, every token is located at a fixed node in S with probability at most $2/n$ (cf. Lemma B.5). Hence we can bound the expectation of Z:

$$\mathbf{E}[Z] \leqslant \left\|x^{(0)}\right\|_1 \cdot \frac{2|S|}{n} \leqslant \mathrm{e}^{-(\log n)^{1-\varepsilon}} \cdot 8 \cdot \mathrm{e}^{\frac{1}{2} \cdot (\log n)^{1-\varepsilon}} = 8 \cdot \mathrm{e}^{-\frac{1}{2}(\log n)^{1-\varepsilon}} < 1.$$

By Lemma 3.3 we have

$$\mathbf{Pr}[Z \geqslant (1+\delta) \cdot \mathbf{E}[Z]] \leqslant \left(\frac{\mathrm{e}^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mathbf{E}[Z]} \leqslant \left(\frac{\mathrm{e}}{\delta}\right)^{\delta \cdot \mathbf{E}[Z]}.$$

Hence choosing $\delta = \frac{15 \cdot (\log n)^{\varepsilon}}{\mathbf{E}[Z]}$ gives

$$\begin{aligned}
\mathbf{Pr}[Z \geqslant 16 \cdot (\log n)^{\varepsilon}] &\leqslant \mathbf{Pr}[Z \geqslant (1+\delta) \cdot \mathbf{E}[Z]] \\
&\leqslant \left(\frac{\mathrm{e} \cdot \mathbf{E}[Z]}{15 \cdot (\log n)^{\varepsilon}}\right)^{15 \cdot (\log n)^{\varepsilon}} \\
&\leqslant \mathrm{e}^{-\frac{1}{2} \cdot (\log n)^{1-\varepsilon} \cdot 15 \cdot (\log n)^{\varepsilon}} \leqslant n^{-7},
\end{aligned}$$

where the last inequality holds for large enough n. This completes the proof of the lemma. ■

After these preparations, we are now able to analyze the drop of the potential function Λ. The following results essentially show that after every $\mathcal{O}\left(\frac{1}{1-\lambda}\right)$ rounds, the potential Λ drops exponentially. First we consider the case where the graph is sparse, i.e., the degree satisfies $d \leqslant \mathrm{e}^{(\log n)^{1/2}}$ and after that we consider the dense case where $d > \mathrm{e}^{(\log n)^{1/2}}$.

Analysis for Sparse Graphs (Random Matching Model). Define for any node $u \in V$ and integer r, $B_r(u) := \{v \in V\colon \operatorname{dist}(u, v) \leqslant r\}$. For any round $t \in \mathbb{N}$, define the event $\mathcal{E}_t$ as

$$\mathcal{E}_t := \bigwedge_{u \in V} \left(\sum_{v \in B_r(u)} x_v^{(t)} \leqslant 16 \cdot (\log n)^{\varepsilon}\right),$$

where $r := (\log n)^{1/3}$. Note that if $\mathcal{E}_t$ happens, then the total number of tokens located at the nodes in each $B_r(u)$ is small.

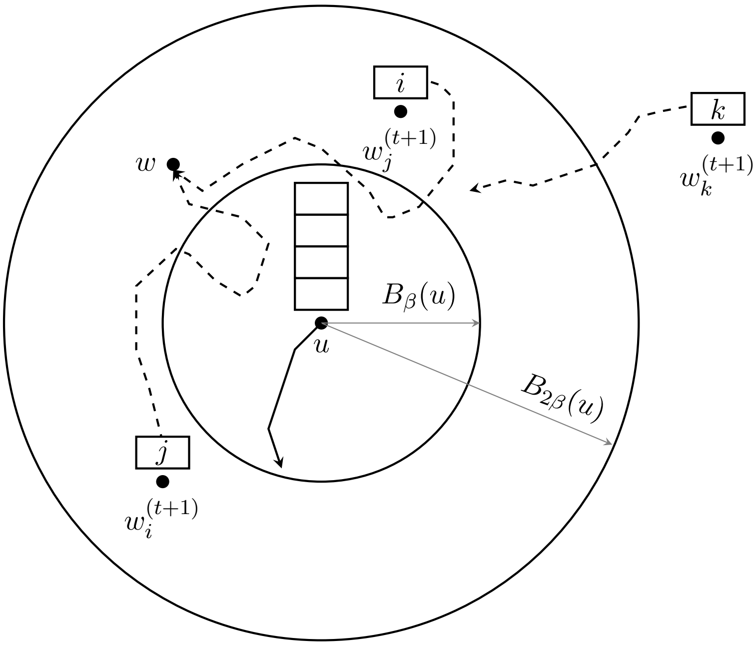


Figure 4: The above diagram illustrates the proof of Lemma 4.15, the key step in proving Theorem 4.3. We show that the load along the canonical path starting from u decreases by at least one within the time-interval $[t+1, t+\beta]$ by analyzing all pairs of tokens i, j within $B_{2\beta}(u)$ and show that none of them is located on the same node w at the end of round $[t+\beta]$. Note that tokens k with $w_k^{(t+1)} \notin B_{2\beta}(u)$ cannot intersect with the canonical path from u within $[t+1, t+\beta]$.

Lemma 4.14. *Consider the random matching model and let G be any d-regular graph with $d \leqslant \mathrm{e}^{(\log n)^{1/2}}$. Let $x^{(0)}$ be a non-negative load vector $x^{(0)}$ with $\|x^{(0)}\|_1 \leqslant n \cdot \mathrm{e}^{-(\log n)^{1-\varepsilon}}$, where $\varepsilon > 0$ is a sufficiently small constant. Then,*

$$\mathbf{Pr}\left[\bigwedge_{t=t_0}^{t_0+n} \mathcal{E}_t\right] \geqslant 1 - 2n^{-1}.$$

Proof. Since $t_0 = \tau_{\text{cont}}(n, n^{-2})$ by definition, the time-interval $[0, t_0]$ is (n, n^{-2})–smoothing with probability at least $1 - n^{-1}$. For the rest of the proof, assume that this happens. Clearly, for any round $t \geqslant t_0$, the time-interval $[0, t]$ is also (n, n^{-2})–smoothing. Since for every $u \in V$ and $\varepsilon > 0$ small enough,

$$\left|B_{(\log n)^{1/3}}(u)\right| \leqslant d^{(\log n)^{1/3}} \leqslant \left(\mathrm{e}^{(\log n)^{1/2}}\right)^{(\log n)^{1/3}} = \mathrm{e}^{(\log n)^{5/6}} < 4 \cdot \mathrm{e}^{\frac{1}{2}\cdot(\log n)^{1-\varepsilon}},$$

we obtain by Lemma 4.13 and the union bound over all n nodes that $\mathbf{Pr}\left[\,\mathcal{E}_t\,\right] \leqslant n^{-6}$. By the union bound over the time-interval $[t_0, t_0+n]$ we have $\mathbf{Pr}\left[\,\bigwedge_{t=t_0}^{t_0+n} \mathcal{E}_t\,\right] \leqslant n^{-5}$, which yields the claim of the lemma. ■

Next we lower bound the potential drop of Λ for load vectors $x^{(t)}$ satisfying $\mathcal{E}_t$.

Lemma 4.15. *Consider the random matching model and let G be any d-regular graph with $d \leqslant \mathrm{e}^{(\log n)^{1/2}}$. Assume that $x^{(t)}$ is any non-negative load vector that satisfies $\mathcal{E}_t$. Then the following two statements hold.*

- *If $\frac{1}{1-\lambda} \leqslant (\log n)^{1/4}$, then*

$$\mathbf{E}\left[\,\Lambda^{(t+\beta)}\,\right] \leqslant \mathrm{e}^{-\Omega((\log n)^{2\varepsilon})} \cdot \Lambda^{(t)},$$

where $\beta := \frac{(\log n)^{2\varepsilon}}{1-\lambda}$ for a sufficiently small constant $\varepsilon > 0$.

- *If* $\frac{1}{1-\lambda} > (\log n)^{1/4}$, *then*

$$\mathbf{E}\left[\Lambda^{(t+\beta)}\right] \leqslant (\log n)^{-\varepsilon} \cdot \Lambda^{(t)},$$

where $\beta := (\log n)^{48\varepsilon}$ *for a sufficiently small constant* $\varepsilon > 0$.

Proof. In both cases we choose ε small enough so that $2 \cdot \beta \leqslant r = (\log n)^{1/3}$.

Case 1: $\frac{1}{1-\lambda} \leqslant (\log n)^{1/4}$. Recall that $\beta = \frac{(\log n)^{2\varepsilon}}{1-\lambda}$. Fix any node $u \in V$ with $x_u^{(t)} \geqslant 2$. Our goal is to prove that the stack of $x_u^{(t)}$ tokens at node $u \in V$ disappears with high probability after β rounds. To this end, we simply consider all tokens in the set $B_{2\beta}(u)$ and bound the probability that any pair of these tokens collide in round $t+\beta$, i.e., share the same location in round $t+\beta$. See Figure 4 for an illustration.

By Corollary 2.4, it holds for any node v that

$$\mathbf{Pr}\left[\left\|\mathbf{M}_{v,\cdot}^{[t+1,t+\beta]}\right\|_2 \leqslant \frac{1}{n} + \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}\right] \geqslant 1 - \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}.$$

Hence for any $\varepsilon < 1/4$ we have $\frac{1}{n} + \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})} = \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}$ and

$$\mathbf{Pr}\left[\left\|\mathbf{M}_{v,\cdot}^{[t+1,t+\beta]}\right\|_\infty \leqslant \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}\right] \geqslant 1 - \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}.$$

Let us now define the following event:

$$\mathcal{A}_u := \bigwedge_{v \in B_{2\beta}(u)} \left(x_v^{(t)} = 0 \bigvee \left\|\mathbf{M}_{v,\cdot}^{[t+1,t+\beta]}\right\|_\infty \leqslant \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}\right),$$

i.e., the event $\mathcal{A}_u$ happens if for all nodes in $B_{2\beta}(u)$ that contain a token in round t the neighborhood within the graph induced by the matchings in the interval $[t+1, t+\beta]$ "expands". Since the load vector $x^{(t)}$ satisfies $\mathcal{E}_t$ by the precondition of the lemma, the total number of tokens in $B_{2\beta}(u)$ is upper bounded by $16 \cdot (\log n)^\varepsilon$. Consequently, the number of nodes with at least one token is also at most $16 \cdot (\log n)^\varepsilon$. Therefore,

$$\mathbf{Pr}\left[\mathcal{A}_u\right] \geqslant 1 - 16 \cdot (\log n)^\varepsilon \cdot \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})} = 1 - \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}. \tag{4.18}$$

Consider any pair of tokens $i, j \in \mathcal{T}$ with $w_i^{(t)} \in B_{2\beta}(u)$ and $w_j^{(t)} \in B_{2\beta}(u)$. We upper bound the probability that token i and j meet in round $t+\beta$ conditioned on $\mathcal{A}_u$ as follows.

$$\begin{aligned}
&\mathbf{Pr}\left[w_i^{(t+\beta)} = w_j^{(t+\beta)} \mid \mathcal{A}_u\right] \\
&= \sum_{w \in V} \mathbf{Pr}\left[w_i^{(t+\beta)} = w \wedge w_j^{(t+\beta)} = w \mid \mathcal{A}_u\right] \\
&\leqslant \sum_{w \in V} \mathbf{Pr}\left[w_i^{(t+\beta)} = w \mid \mathcal{A}_u\right] \cdot \mathbf{Pr}\left[w_j^{(t+\beta)} = w \mid \mathcal{A}_u\right] \\
&\leqslant \max_{v \in V} \mathbf{Pr}\left[w_i^{(t+\beta)} = v \mid \mathcal{A}_u\right] \cdot \sum_{w \in V} \mathbf{Pr}\left[w_j^{(t+\beta)} = w \mid \mathcal{A}_u\right] \\
&\leqslant \max_{v \in V} \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})} \cdot 1 = \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})},
\end{aligned} \tag{4.19}$$

where the first inequality follows by Lemma 3.2 and the last inequality follows from the definition

of $\mathcal{A}_u$. Recall that since $x^{(t)}$ satisfies $\mathcal{E}_t$, there are at most $16 \cdot (\log n)^{\varepsilon}$ tokens in $B_{2\beta}(u)$. Hence,

$$\begin{aligned}
&\mathbf{Pr}\left[\bigvee_{\substack{i\in\mathcal{T}\colon w_i^{(t)}\in B_{2\beta}(u)\\ j\in\mathcal{T}\colon w_j^{(t)}\in B_{2\beta}(u)}} \left(w_i^{(t+\beta)} = w_j^{(t+\beta)}\right)\right]\\
&\leqslant \mathbf{Pr}\left[\bigvee_{\substack{i\in\mathcal{T}\colon w_i^{(t)}\in B_{2\beta}(u)\\ j\in\mathcal{T}\colon w_j^{(t)}\in B_{2\beta}(u)}} \left(w_i^{(t+\beta)} = w_j^{(t+\beta)}\right) \,\middle|\, \mathcal{A}_u\right] + \mathbf{Pr}\left[\neg\mathcal{A}_u\right]\\
&\leqslant 256(\log n)^{2\varepsilon} \cdot \max_{\substack{i\in\mathcal{T}\colon w_i^{(t)}\in B_{2\beta}(u)\\ j\in\mathcal{T}\colon w_j^{(t)}\in B_{2\beta}(u)}} \mathbf{Pr}\left[w_i^{(t+\beta)} = w \wedge w_j^{(t+\beta)} = w \mid \mathcal{A}_u\right] + \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}\\
&\leqslant 256(\log n)^{2\varepsilon} \cdot \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})} + \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})} = \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})},
\end{aligned}$$

where the last inequality follows from (4.19). Note that in case there are no tokens in $B_{2\beta}(u)$ which share the same node in round $t+\beta$, then if we follow the canonical path which starts from u in round t, there is at least one round $t' \in [t+1, t+\beta]$ in which the load on the canonical path of u from round t is at least $x_u^{(t)}$ and is reduced by one in round $t'+1$. Since two canonical paths which meet in a certain round t' cannot both reduce their value, we obtain by the second statement of Lemma 4.12 that

$$\begin{aligned}
\mathbf{E}\left[\Lambda^{(t+\beta)} - \Lambda^{(t)}\right] &\geqslant \sum_{u\colon x_u^{(t)}\geqslant 2} \Lambda_u^{(t)} \cdot \left(1 - \frac{2}{\mathrm{e}^{\frac{1}{8}(\log n)^{1-\varepsilon}}}\right) \cdot \left(1 - \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}\right)\\
&\geqslant \sum_{u\colon x_u^{(t)}\geqslant 2} \Lambda_u^{(t)} \cdot \left(1 - \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}\right) = \left(1 - \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})}\right) \cdot \Lambda^{(t)},
\end{aligned}$$

where the second inequality holds for $\varepsilon < 1/3$. Rearranging gives

$$\mathbf{E}\left[\Lambda^{(t+\beta)}\right] \leqslant \mathrm{e}^{-\Theta((\log n)^{2\varepsilon})} \cdot \Lambda^{(t)},$$

which finishes the first case.

Case 2: $\frac{1}{1-\lambda} > (\log n)^{1/4}$. Now we proceed similarly as in the first case, but here we have $\beta = (\log n)^{48\varepsilon}$, where $\varepsilon > 0$ is a sufficiently small constant. By Lemma B.4, it holds that for any node $u \in V$:

$$\mathbf{Pr}\left[\left\|\mathbf{M}_{u,\cdot}^{[t+1,t+\beta]}\right\|_2^2 \leqslant (\log n)^{-8\varepsilon}\right] \geqslant 1 - \mathrm{e}^{-(\log n)^{32\varepsilon}}$$

We redefine $\mathcal{A}_u$ as follows:

$$\mathcal{A}_u := \bigwedge_{v\in B_{2\beta}(u)} \left(x_v^{(t)} = 0 \bigvee \left\|\mathbf{M}_{v,\cdot}^{[t+1,t+\beta]}\right\|_\infty \leqslant (\log n)^{-4\varepsilon}\right).$$

As $x^{(t)}$ satisfies $\mathcal{E}_t$ by the precondition of the lemma, the set $B_{2\beta}(u)$ contains at most $16 \cdot (\log n)^{\varepsilon}$ tokens in total in round t and this event is independent of the random choices for the matchings in the rounds after t. Similar to (4.18) we have

$$\mathbf{Pr}\left[\mathcal{A}_u\right] \geqslant 1 - 16 \cdot (\log n)^{\varepsilon} \cdot \mathrm{e}^{-(\log n)^{32\varepsilon}} = 1 - \mathrm{e}^{-\Theta((\log n)^{32\varepsilon})}.$$

As in the first case, we conclude that

$$\begin{aligned}
&\mathbf{Pr}\left[\bigvee_{\substack{i\in\mathcal{T}\colon w_i^{(t)}\in B_{2\beta}(u)\\ j\in\mathcal{T}\colon w_j^{(t)}\in B_{2\beta}(u)}}\left(w_i^{(t+\beta)}=w_j^{(t+\beta)}\right)\right]\\
&\leqslant \mathbf{Pr}\left[\bigvee_{\substack{i\in\mathcal{T}\colon w_i^{(t)}\in B_{2\beta}(u)\\ j\in\mathcal{T}\colon w_j^{(t)}\in B_{2\beta}(u)}}\left(w_i^{(t+\beta)}=w_j^{(t+\beta)}\right)\,\middle|\,\mathcal{A}_u\right]+\mathbf{Pr}\left[\neg\mathcal{A}_u\right]\\
&\leqslant 256\cdot(\log n)^{2\varepsilon}\cdot(\log n)^{-4\varepsilon}+\mathrm{e}^{-\Theta((\log n)^{32\varepsilon})}=o((\log n)^{-\varepsilon}),
\end{aligned}$$

and as in the first case we conclude that $\mathbf{E}\left[\Lambda^{(t+\beta)}\right]\leqslant(\log n)^{-\varepsilon}\cdot\Lambda^{(t)}$. ■

Analysis for Dense Graphs (Random Matching Model). We now consider the dense case where the degree of the graph satisfies $d\geqslant \mathrm{e}^{(\log n)^{1/2}}$. This case is easier than the sparse case, as in this case the average load around every node is smaller than $1/2$ for all nodes and all rounds with high probability. Hence most of the neighbors of each node have zero tokens which implies that as long as the maximum load is larger than 1, there is an expected exponential drop of the potential function within a single round.

To formalize this, let us define for any round $t\in\mathbb{N}$ the following event:

$$\mathcal{F}_t:=\bigwedge_{u\in V}\left(\sum_{v\in N(u)}x_v^{(t)}\leqslant\frac{d}{2}\right).$$

Similar to Lemma 4.14, we now prove the following:

Lemma 4.16. *Consider the random matching model and let G be any d-regular graph with $d>\mathrm{e}^{(\log n)^{1/2}}$. Let $x^{(0)}$ be a non-negative load vector $x^{(0)}$ with $\|x^{(0)}\|_1\leqslant n\cdot\mathrm{e}^{-(\log n)^{1-\varepsilon}}$, where $\varepsilon>0$ is a sufficiently small constant. Then,*

$$\mathbf{Pr}\left[\bigwedge_{t=t_0}^{t_0+n}\mathcal{F}_t\right]\geqslant 1-2n^{-1}.$$

Proof. Since $t_0=\tau_{\mathrm{cont}}(n,n^{-2})$ by definition, the time-interval $[0,t_0]$ is (n,n^{-2})–smoothing with probability at least $1-n^{-1}$. For the rest of the proof, assume that this happens. Clearly, this also implies that for any round $t\geqslant t_0$, the time-interval $[0,t]$ is also (n,n^{-2})–smoothing. Let us first lower bound $\mathbf{Pr}\left[\mathcal{F}_t\right]$ for any fixed $t\geqslant t_0$. Consider any round $t\geqslant t_0$ and fix a node $u\in V$. Let

$$Z:=\sum_{v\in N(u)}x_v^{(t)}.$$

Since $[0,t]$ is (n,n^{-2})–smoothing, every token is located at a fixed node in S with probability at most $2/n$ (cf. Lemma B.5). Therefore,

$$\mathbf{E}\left[Z\right]=|N(u)|\cdot\|x^{(0)}\|_1\cdot\frac{2}{n}\leqslant 2d\cdot\mathrm{e}^{-(\log n)^{1-\varepsilon}}.$$

By Lemma 3.3,

$$\mathbf{Pr}\left[Z\geqslant(1+\delta)\mathbf{E}\left[Z\right]\right]\leqslant\left(\frac{\mathrm{e}^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mathbf{E}[Z]}\leqslant\delta^{-\delta\mathbf{E}[Z]/2},$$

for any $\delta > 9$. Choosing $\delta = d/(4\mathbf{E}[\,Z\,])$ yields,

$$\mathbf{Pr}[\,Z > d/2\,] \leqslant \delta^{-\delta\mathbf{E}[\,Z\,]/2} = \left(\frac{d}{4\mathbf{E}[\,Z\,]}\right)^{-d/8} \leqslant n^{-\omega(1)}.$$

By the union bound,

$$\mathbf{Pr}\left[\bigwedge_{t=t_0}^{t_0+n} \mathcal{F}_t\right] = 1 - n^{-\omega(1)},$$

and recalling that with probability at least $1 - n^{-1}$, all the intervals $[0,t], t \geqslant t_0$ are (n, n^{-2})–smoothing, completes the proof. ■

The next lemma is similar to Lemma 4.15 from the sparse graph case.

Lemma 4.17. *Consider the random matching model and let G be any d-regular graph with $d \geqslant \mathrm{e}^{(\log n)^{1/2}}$. Assume that $x^{(t)}$ is any non-negative load vector that satisfies $\mathcal{E}_t$. Then there is a constant $\gamma \in (0,1)$ such that $\mathbf{E}\left[\,\Lambda^{(t+1)}\,\right] \leqslant (1-\gamma)\cdot\Lambda^{(t)}$.*

Proof. Note that

$$\begin{aligned}\mathbf{E}\left[\Lambda^{(t)} - \Lambda^{(t+1)}\right] &\geqslant \sum_{\substack{u \in V:\\ x_u^{(t)} \geqslant 2}} \sum_{\substack{v \in N(u):\\ x_v^{(t)} = 0}} \mathbf{Pr}\left[\{u,v\} \in \mathbf{M}^{(t)}\right] \cdot \left(\mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}\cdot x_u^{(t)}} \cdot \left(1 - \frac{2}{\mathrm{e}^{\frac{1}{8}\cdot(\log n)^{1-\varepsilon}}}\right)\right)\\ &\geqslant \frac{1}{2}\cdot\sum_{\substack{u \in V:\\ x_u^{(t)} \geqslant 2}} \sum_{\substack{v \in N(u):\\ x_v^{(t)} = 0}} \Omega\left(\frac{1}{d}\right)\cdot \mathrm{e}^{\frac{1}{8}(\log n)^{1-\varepsilon}\cdot x_u^{(t)}}.\end{aligned}$$

Since $x^{(t)}$ satisfies $\mathcal{E}_t$, we know that for every node $u \in V$, at least half of the neighbors $v \in N(u)$ satisfy $x_v^{(t)} = 0$. Therefore,

$$\mathbf{E}\left[\Lambda^{(t)} - \Lambda^{(t+1)}\right] \geqslant \Omega(1)\cdot \sum_{\substack{u \in V:\\ x_u^{(t)} \geqslant 2}} \mathrm{e}^{\frac{1}{8}(\log n)^{1-\varepsilon}\cdot x_u^{(t)}} = \gamma\cdot\Lambda^{(t)}$$

for a constant $\gamma \in (0,1)$. Hence $\mathbf{E}\left[\,\Lambda^{(t+1)}\,\right] \leqslant (1-\gamma)\cdot\Lambda^{(t)}$. ■

Finally, we are able to prove Theorem 4.3.

Proof of Theorem 4.3. We give the proof for the random matching model at first (where d denotes the degree of the graph) and consider the balancing circuit model at the end of the proof (where d represents the number of matchings which are applied periodically). Corresponding to Lemma 4.15 and Lemma 4.17, we divide our proof into four different cases.

Case 1: $d \leqslant \mathrm{e}^{(\log n)^{1/2}}$ and $\frac{1}{1-\lambda} \leqslant (\log n)^{1/4}$. By the first statement of Lemma 4.12, it holds with probability $1-2n^{-1}$ that $\Lambda^{(t_0)} \leqslant 9n^2$. Let $\mathcal{E} := \bigwedge_{t=t_0}^{t_0+n}\mathcal{E}_t$, where $t_0 := \tau_{\mathrm{cont}}(n, n^{-2})$. By Lemma 4.14,

$$\mathbf{Pr}[\,\mathcal{E}\,] \geqslant 1 - 2n^{-1}.$$

Moreover, by Lemma 4.15 with $\beta = \frac{(\log n)^{2\varepsilon}}{1-\lambda}$, we have

$$\mathbf{E}\left[\Lambda^{(t+\beta)} \mid \mathcal{E}_t\right] \leqslant \mathrm{e}^{-\Omega((\log n)^{2\varepsilon})}\cdot\Lambda^{(t)},$$

where $\varepsilon > 0$ is a sufficiently small constant.

As in (4.14), we obtain that

$$\mathbf{E}\left[\Lambda^{(t+\beta)} \mid \mathcal{E}\right] \leqslant \mathrm{e}^{-\Omega((\log n)^{2\varepsilon})}\cdot\Lambda^{(t)}.$$

Iterating this inequality $\tau = C \cdot \frac{\log n}{\beta \cdot (1-\lambda)}$ times starting from round t_0, and using $\Lambda^{(t_0)} \leqslant 9n^2$ gives

$$\mathbf{E}\left[\Lambda^{(t_0+\tau\cdot\beta)} \mid \mathcal{E} \right] \leqslant \mathrm{e}^{-\tau\cdot\Omega((\log n)^{2\varepsilon})} \cdot \Lambda^{(t_0)} \leqslant n^{-1},$$

if $C > 0$ is a sufficiently large constant. Since $\tau \cdot \beta = \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$, Markov's inequality implies

$$\mathbf{Pr}\left[\Lambda^{(t_0+\tau\cdot\beta)} \geqslant 1 \right] \leqslant n^{-1}.$$

Since $\Lambda^{(t_0+\tau\cdot\beta)} < 1$ implies $\Lambda^{(t_0+\tau\cdot\beta)} = 0$ which is equivalent to a maximum load of one, the proof of Case 1 is complete.

Case 2: $d \leqslant \mathrm{e}^{(\log n)^{1/2}}$ and $\frac{1}{1-\lambda} > (\log n)^{1/4}$. The proof of this case is very similar. By the second statement of Lemma 4.15, we have

$$\mathbf{E}\left[\Lambda^{(t+\beta)} \mid \mathcal{E}_t \right] \leqslant (\log n)^{-\varepsilon} \cdot \Lambda^{(t)},$$

where $\beta = (\log n)^{48\varepsilon}$ and $\varepsilon > 0$ is a sufficiently small constant. As the analysis in (4.14), we obtain that

$$\mathbf{E}\left[\Lambda^{(t+\beta)} \mid \mathcal{E} \right] \leqslant 2 \cdot (\log n)^{-\varepsilon} \cdot \Lambda^{(t)}.$$

Iterating this inequality $\tau := \log n$ times and using $\Lambda^{(t_0)} \leqslant 9n^2$, gives

$$\mathbf{E}\left[\Lambda^{(t_0+\tau\cdot\beta)} \mid \mathcal{E} \right] \leqslant 2 \cdot (\log n)^{-\varepsilon\cdot\log n} \cdot \Lambda^{(t_0)} = n^{-\omega(1)}.$$

Note that $\tau \cdot \beta = (\log n)^{48\varepsilon} \cdot \log n$. Since $\frac{1}{1-\lambda} \geqslant (\log n)^{1/4}$, $\tau \cdot \beta = \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$ if $\varepsilon < \frac{1}{192}$. The rest of the proof is exactly the same as the proof in Case 1.

Case 3: $d > \mathrm{e}^{(\log n)^{1/2}}$. The proof of this case is the same as Case 1 and 2. In fact, it is even slightly simpler, because Lemma 4.17 implies an exponential drop of the potential Λ in a single round.

Case 4: Balancing Circuit Model. Finally, we consider the balancing circuit model with a sequence of $d = \mathcal{O}(1)$ matchings. Since d is a constant, it holds for every node $u \in V$ that at most d neighbors of u in G appear in the matching matrices $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(d)}$. For this reason, we assume w.l.o.g. that the underlying graph G has bounded maximum degree. Moreover, we can apply Corollary B.2 and Lemma B.3 to conclude that for all nodes $u \in V$ and any $\beta \in \mathbb{N}$,

$$\left\| \mathbf{M}^{\beta}_{u,\cdot} \right\|_2^2 \leqslant \frac{1}{n} + \min\left\{ \mathcal{O}(\beta^{-1/2}), \lambda(\mathbf{M})^{2\beta} \right\}. \tag{4.20}$$

Therefore, the same statements as in Lemma 4.13, Lemma 4.14 and Lemma 4.15 hold for the balancing circuit model with the only difference that every round in the random matching model corresponds to d consecutive rounds in the balancing circuit model. Consequently, the analysis of the balancing circuit model is the same as the analysis of the random matching model (Case 1 and Case 2). ■

5 The Diffusion Model

In the diffusion model, the (continuous) load vector $\xi^t \in \mathbb{R}^n$ in round $t \geqslant 1$ is given by the recursion $\xi^{(t)} = \xi^{(t-1)}\mathbf{P}$, where for any $\gamma \geqslant 1$ the *diffusion matrix* $\mathbf{P} = \mathbf{P}(\gamma)$ of graph G is defined as follows: For $u \in V$, $\mathbf{P}_{u,v} = \frac{1}{\gamma\Delta}$ if $\{u,v\} \in E$, $\mathbf{P}_{u,v} = 1 - \frac{d(u)}{\gamma\Delta}$ if $u = v$, and $\mathbf{P}_{u,v} = 0$ otherwise. Hence,

$$\xi^{(t)}_u = \xi^{(t-1)}_u + \sum_{v:\{u,v\}\in E} \frac{\xi^{(t-1)}_v - \xi^{(t-1)}_u}{\gamma\Delta}.$$

Common choices are $\gamma = 2$ (resulting in a loop probability at least 1/2) or $\gamma = 1+1/\Delta$, both ensuring convergence also on bipartite graphs. Let $\lambda_1(\mathbf{P}) \geqslant \ldots \geqslant \lambda_n(\mathbf{P})$ be the eigenvalues of $\mathbf{P}$ and $\lambda(\mathbf{P}) := \max\{|\lambda_2(\mathbf{P})|, |\lambda_n(\mathbf{P})|\}$. Note that $\lambda(\mathbf{P})$ depends on γ. Let $\mathbf{P}^t$ be the t-th power of $\mathbf{P}$ and $\mathbf{P}^0$ be the n by n identity matrix.

As for the matching model, there is a natural upper bound on the convergence in terms of the spectral gap of $\mathbf{P}$.

Theorem 5.1 ([35, Thm. 1])**.** *Let G be any graph and consider the continuous case. Then for any $\varepsilon > 0$, the discrepancy is at most ε after $\frac{2}{1-\lambda(\mathbf{P})} \cdot \log(\frac{Kn^2}{\varepsilon})$ rounds for any initial load vector with discrepancy at most K.*

5.1 The Discrete Case

We study two natural protocols in the discrete case of the diffusion model. One is the vertex-based protocol [8], where excess tokens are allocated by vertices. The second protocol we study is the edge-based protocol [18], where every edge performs an independent randomized rounding.

The *vertex-based protocol* from [8] works for d-regular graphs as follows. In round t, every node u sends first $\lfloor \frac{x_u^{(t-1)}}{d+1} \rfloor$ tokens to each neighbor and keeps the same amount of tokens for itself. After that, the remaining $x_u^{(t-1)} - (d+1) \cdot \lfloor \frac{x_u^{(t-1)}}{d+1} \rfloor$ tokens at node u are randomly distributed (without replacement) among node u and its d neighbors. This corresponds to a diffusion matrix $\mathbf{P}$ with $\gamma = 1 + 1/d$.

Consider now the edge-based protocol [18], where the load sent along each edge is obtained by randomly rounding the flow that would be sent in the continuous case to a nearest integer. Let $x^{(0)} = \xi^{(0)} \in \mathbb{Z}^n$. As in the matching model, we can derive the following expression for the deviation between the discrete and continuous model at some node w in round t:

$$x_w^{(t)} - \xi_w^{(t)} = \sum_{s=1}^{t} \sum_{u \in V} e_u^{(s)} \mathbf{P}_{u,w}^{t-s} = \sum_{s=1}^{t} \sum_{u \in V} \sum_{v:\ \{u,v\} \in E} e_{u,v}^{(s)} \mathbf{P}_{u,w}^{t-s}$$
$$= \sum_{s=1}^{t} \sum_{[u:v] \in E} e_{u,v}^{(s)} \left(\mathbf{P}_{u,w}^{t-s} - \mathbf{P}_{v,w}^{t-s} \right), \tag{5.1}$$

where $e_{u,v}^{(s)}$ is the rounding error for each edge $[u:v] \in E$ in round s defined by

$$e_{u,v}^{(s)} = \begin{cases} \left\lceil \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta} \right\rceil - \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta} & \text{w. p. } \quad \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta} - \left\lfloor \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta} \right\rfloor \\ \left\lfloor \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta} \right\rfloor - \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta} & \text{w. p. } \quad \left\lceil \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta} \right\rceil - \dfrac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta}. \end{cases}$$

Moreover if $\frac{\xi_v^{(s-1)} - \xi_u^{(s-1)}}{\gamma\Delta}$ is an integer, then $e_{u,v}^{(s)} = 0$. By definition, we have $\mathbf{E}\left[e_{u,v}^{(s)} \right] = 0$. Further, for any set of different (not necessarily disjoint) edges, their rounding errors within the same round are mutually independent.

5.2 Local Divergence and Discrepancy

Based on the deviation between the discrete and continuous case (5.1), we now define the (refined) local p-divergence in the diffusion model.

Definition 5.2 (Local p-Divergence for Diffusion [8, 17, 35])**.** *For any graph G and $p \in \mathbb{Z}_+$ the local p-divergence is*

$$\Psi_p(\mathbf{P}) = \max_{w \in V} \left(\sum_{t=0}^{\infty} \sum_{[u:v] \in E} \left| \mathbf{P}^t_{w,u} - \mathbf{P}^t_{w,v} \right|^p \right)^{1/p},$$

and the refined local p-divergence is

$$\Upsilon_p(\mathbf{P}) = \max_{w \in V} \left(\frac{1}{2} \sum_{t=0}^{\infty} \sum_{u \in V} \max_{v \in N(u)} \left| \mathbf{P}^t_{w,u} - \mathbf{P}^t_{w,v} \right|^p \right)^{1/p}.$$

Clearly, $\Upsilon_p(\mathbf{P}) \leqslant \Psi_p(\mathbf{P})$. We now present our bounds on the (refined) local 2-divergence.

Theorem 5.3. *For any graph G and any $\gamma > 1$ not necessarily constant, it holds that*

$$\Upsilon_2(\mathbf{P}) \leqslant \Psi_2(\mathbf{P}) \leqslant \sqrt{\frac{\gamma \cdot \Delta}{2 - 2/\gamma}}.$$

Moreover, for any $\gamma > 0$, we have $\Psi_2(\mathbf{P}) \geqslant \sqrt{\Delta}$ and

$$\Upsilon_2(\mathbf{P}) \geqslant \sqrt{\frac{1+\Delta}{2}}.$$

The upper bound on $\Psi_2(\mathbf{P})$ is minimized for $\gamma = 2$ and becomes in that case $\sqrt{2 \cdot \Delta}$. This result significantly improves over the previous bounds in [8], which all depend on the spectral gap $1 - \lambda(\mathbf{P})$ or are restricted to special networks. The analysis of the edge-based algorithm in [18] did not use $\Psi_2(\mathbf{P})$ or $\Upsilon_2(\mathbf{P})$, but their bound on the discrepancy also includes the spectral gap.

Proof of Theorem 5.3. Our proof uses a similar approach based on the same potential function as in [7, Lemma 1]. However, we have to perform a more precise analysis to handle the case where γ is very close to 1, i.e. $\gamma = 1 + 1/d$ which corresponds to the vertex-based protocol. By contrast, the proof in [7, Lemma 1] is based on a sequential exposure of the edges and only works if $\gamma \geqslant 4$.

Fix any node $w \in V$. Define the potential function in round t by

$$\Phi^{(t)} := \sum_{u \in V} \left(\mathbf{P}^t_{w,u} - \frac{1}{n} \right)^2.$$

Clearly, $\Phi^{(0)} = 1 \cdot (1 - \frac{1}{n})^2 + (n-1) \cdot (\frac{1}{n})^2 = 1 - \frac{1}{n}$.

We first prove the upper bound on $\Phi^{(t)}$. Let $y_u := \mathbf{P}^{t-1}_{w,u}$. By the definition of the diffusion model, we have

$$\begin{aligned}
\Phi^{(t)} &= \sum_{u \in V} \left(\mathbf{P}^t_{w,u} - \frac{1}{n} \right)^2 \\
&= \sum_{u \in V} \left(\left(1 - \frac{d(u)}{\gamma\Delta}\right) \cdot y_u + \left(\sum_{v \in N(u)} \frac{y_v}{\gamma\Delta} \right) - \frac{1}{n} \right)^2 \\
&= \sum_{u \in V} \left(\left(1 - \frac{d(u)}{\gamma\Delta}\right) \cdot \left(y_u - \frac{1}{n}\right) + \sum_{v \in N(u)} \frac{1}{\gamma\Delta} \left(y_v - \frac{1}{n}\right) \right)^2 \\
&= \sum_{u \in V} \left[\sum_{v \in N(u)} \frac{1}{d(u)} \cdot \left(\left(1 - \frac{d(u)}{\gamma\Delta}\right) \cdot \left(y_u - \frac{1}{n}\right) + \frac{d(u)}{\gamma\Delta} \left(y_v - \frac{1}{n}\right) \right) \right]^2.
\end{aligned}$$

Using the notation $\mathbf{E}_{v\in N(u)}\left[\,X(v)\,\right]$ for the expection of random variable $X(v)$ where $v \in N(u)$ is chosen uniformly at random, we can rewrite the above expression and upper bound it using Jensen's inequality:

$$
\begin{aligned}
\Phi^{(t)} &= \sum_{u\in V}\left(\mathbf{E}_{v\in N(u)}\left[\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right]\right)^2\\
&\leqslant \sum_{u\in V}\mathbf{E}_{v\in N(u)}\left[\left(\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)^2\right]\\
&= \sum_{u\in V}\sum_{v\in N(u)}\frac{1}{d(u)}\cdot\left(\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)^2\\
&= \sum_{[u:v]\in E}\left\{\frac{1}{d(u)}\cdot\left(\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)^2\right.\\
&\qquad\left.+\frac{1}{d(v)}\cdot\left(\left(1-\frac{d(v)}{\gamma\Delta}\right)\cdot\left(y_v-\frac{1}{n}\right)+\frac{d(v)}{\gamma\Delta}\cdot\left(y_u-\frac{1}{n}\right)\right)^2\right\}.
\end{aligned}
$$

Note that

$$
\Phi^{(t-1)} = \sum_{u\in V}\left(y_u-\frac{1}{n}\right)^2 = \sum_{[u:v]\in E}\left\{\frac{1}{d(u)}\cdot\left(y_u-\frac{1}{n}\right)^2+\frac{1}{d(v)}\cdot\left(y_v-\frac{1}{n}\right)^2\right\},
$$

and using the upper bound on $\Phi^{(t)}$ from above, we obtain

$$
\begin{aligned}
&\Phi^{(t-1)}-\Phi^{(t)}\\
&\geqslant \sum_{[u:v]\in E}\left\{\frac{1}{d(u)}\cdot\left(y_u-\frac{1}{n}\right)^2+\frac{1}{d(v)}\cdot\left(y_v-\frac{1}{n}\right)^2\right\}\\
&\quad-\sum_{[u:v]\in E}\left\{\frac{1}{d(u)}\cdot\left(\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)^2\right.\\
&\qquad\qquad\left.+\frac{1}{d(v)}\cdot\left(\left(1-\frac{d(v)}{\gamma\Delta}\right)\cdot\left(y_v-\frac{1}{n}\right)+\frac{d(v)}{\gamma\Delta}\cdot\left(y_u-\frac{1}{n}\right)\right)^2\right\}\\
&= \sum_{[u:v]\in E}\left\{\underbrace{\frac{1}{d(u)}\cdot\left(y_u-\frac{1}{n}\right)^2-\frac{1}{d(u)}\cdot\left(\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)^2}_{=:A}\right.\\
&\qquad\left.+\underbrace{\frac{1}{d(v)}\cdot\left(y_v-\frac{1}{n}\right)^2-\frac{1}{d(v)}\cdot\left(\left(1-\frac{d(v)}{\gamma\Delta}\right)\cdot\left(y_v-\frac{1}{n}\right)+\frac{d(v)}{\gamma\Delta}\cdot\left(y_u-\frac{1}{n}\right)\right)^2}_{=:B}\right\}
\end{aligned}
$$

Let us compute A at first.

$$
\begin{aligned}
A &= \frac{1}{d(u)}\cdot\left[\left(y_u-\frac{1}{n}\right)^2-\left(\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)^2\right]\\
&= \frac{1}{d(u)}\cdot\left(y_u-\frac{1}{n}+\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\left(y_v-\frac{1}{n}\right)\right)\\
&\quad\cdot\left(y_u-\frac{1}{n}-\left(1-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)-\frac{d(u)}{\gamma\Delta}\left(y_v-\frac{1}{n}\right)\right),
\end{aligned}
$$

where the last equality follows from $p^2 - q^2 = (p+q)\cdot(p-q)$. Further,

$$\begin{aligned}
A &= \frac{1}{d(u)}\left(\left(2-\frac{d(u)}{\gamma\Delta}\right)\cdot\left(y_u-\frac{1}{n}\right)+\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)\cdot\left(\frac{d(u)}{\gamma\Delta}\cdot\left(y_u-\frac{1}{n}\right)-\frac{d(u)}{\gamma\Delta}\cdot\left(y_v-\frac{1}{n}\right)\right)\\
&= \frac{1}{d(u)}\cdot\left(2y_u-\frac{2}{n}-\frac{d(u)}{\gamma\Delta}(y_u-y_v)\right)\cdot\frac{d(u)}{\gamma\Delta}\cdot(y_u-y_v)\\
&= \frac{1}{\gamma\Delta}\cdot\left(y_u-y_v\right)\cdot\left(2y_u-\frac{2}{n}-\frac{d(u)}{\gamma\Delta}\cdot(y_u-y_v)\right).
\end{aligned}$$

Similarly, we get

$$B = \frac{1}{\gamma\Delta}\cdot\left(y_v-y_u\right)\cdot\left(2y_v-\frac{2}{n}-\frac{d(v)}{\gamma\Delta}\cdot(y_v-y_u)\right),$$

and thus

$$\begin{aligned}
A+B &= \frac{1}{\gamma\Delta}\cdot\left(y_u-y_v\right)\cdot\left(2y_u-\frac{2}{n}-\frac{d(u)}{\gamma\Delta}\cdot\left(y_u-y_v\right)-2y_v+\frac{2}{n}+\frac{d(v)}{\gamma\Delta}\cdot(y_v-y_u)\right)\\
&= \frac{1}{\gamma\Delta}\cdot\left(y_u-y_v\right)\cdot\left(y_u-y_v\right)\cdot\left(2-\frac{d(u)+d(v)}{\gamma\Delta}\right)\\
&\geqslant \frac{1}{\gamma\Delta}\cdot\left(y_u-y_v\right)^2\cdot\left(2-\frac{2\Delta}{\gamma\Delta}\right)\\
&= \frac{1}{\gamma\Delta}\cdot\left(y_u-y_v\right)^2\cdot\left(2-\frac{2}{\gamma}\right)\ .
\end{aligned}$$

Therefore,

$$\Phi^{(t-1)}-\Phi^{(t)} \geqslant \frac{1}{\gamma\Delta}\cdot\left(2-\frac{2}{\gamma}\right)\cdot\sum_{[u:v]\in E}\left(y_u-y_v\right)^2,$$

i.e.

$$\sum_{[u:v]\in E}\left(\mathbf{P}^{t-1}_{w,u}-\mathbf{P}^{t-1}_{w,v}\right)^2 \leqslant \frac{\gamma\Delta}{2-2/\gamma}\left(\Phi^{(t-1)}-\Phi^{(t)}\right)\ .$$

Finally, summing over all rounds gives

$$\sum_{t=1}^{\infty}\sum_{[u:v]\in E}\left(\mathbf{P}^{t-1}_{w,u}-\mathbf{P}^{t-1}_{w,v}\right)^2 \leqslant \frac{\gamma\Delta}{2-2/\gamma}\sum_{t=1}^{\infty}\left(\Phi^{(t-1)}-\Phi^{(t)}\right) \leqslant \frac{\gamma\Delta}{2-2/\gamma}\cdot\Phi^{(0)} = \frac{\gamma\Delta}{2-2/\gamma}\left(1-\frac{1}{n}\right)$$

and $\Psi_2(\mathbf{P}) < \sqrt{\frac{\gamma\cdot\Delta}{2-2/\gamma}}$. For the lower bound on $\Psi_2(\mathbf{P})$, we consider a node $w\in V$ with $d(w)=\Delta$ to obtain that

$$\Psi_2(\mathbf{P}) \geqslant \sqrt{\sum_{[u:v]\in E}\left(\mathbf{P}^{0}_{w,u}-\mathbf{P}^{0}_{w,v}\right)^2} \geqslant \sqrt{\Delta\cdot(1-0)^2} = \sqrt{\Delta}.$$

In the same way, we prove the lower bound on $\Upsilon_2(\mathbf{P})$:

$$\Upsilon_2(\mathbf{P}) \geqslant \sqrt{\frac{1}{2}\sum_{u\in V}\max_{v\in N(u)}\left(\mathbf{P}^{0}_{w,u}-\mathbf{P}^{0}_{w,v}\right)^2} = \sqrt{\frac{1}{2}(\Delta+1)\cdot 1}\ . \qquad \blacksquare$$

Lemma 5.4. *Consider the edge-based diffusion model. Fix two rounds $t_1 < t_2$ and the load vector $x^{(t_1)}$ at the end of round t_1. For any family of non-negative numbers $g^{(s)}_{u,v}$ ($[u:v]\in$*

E, $t_1{+}1 \leqslant s \leqslant t_2$), define the random variable $Z := \sum_{s=t_1+1}^{t_2} \sum_{[u:v]\in E} g_{u,v}^{(s)} \cdot e_{u,v}^{(s)}$. Then $\mathbf{E}\left[\, Z \,\right] = 0$ and for any $\delta > 0$ it holds that

$$\mathbf{Pr}\left[\, |Z - \mathbf{E}\left[\, Z \,\right]| \geqslant \delta \,\right] \leqslant 2\exp\left(- \frac{\delta^2}{8 \sum_{s=t_1+1}^{t_2} \sum_{[u:v]\in E} \left(g_{u,v}^{(s)}\right)^2} \right).$$

The proof of Lemma 5.4 is the same as Lemma 2.12, except that the inner sum runs over all edges of the graph and we have to use the slightly weaker inequality $|e_{u,v}^{(s)}| < 1$ instead of $|e_{u,v}^{(s)}| \leqslant 1/2$, which results into an extra factor of 4 in the denominator.

We now use the above machinery to derive upper bounds on the discrepancy for the edge-based and vertex-based protocol.

Theorem 5.5. *Consider the edge-based protocol on an arbitrary graph G where $\gamma > 1$ is any constant. Then:*

- *For any round t,*

$$\mathbf{Pr}\left[\max_{w\in V} \left| x_w^{(t)} - \xi_w^{(t)} \right| = \mathcal{O}\left(\sqrt{\Delta \log n} \right) \right] \geqslant 1 - 2n^{-1}.$$

- *After $\mathcal{O}\left(\frac{\log(Kn)}{1-\lambda(\mathbf{P})}\right)$ rounds, the discrepancy is at most $\mathcal{O}(\sqrt{\Delta \log n})$ w. p. at least $1 - n^{-1}$.*

Moreover, consider the vertex-based protocol on a d-regular graph G. Then:

- *For any round t,*

$$\mathbf{Pr}\left[\max_{w\in V} \left| x_w^{(t)} - \xi_w^{(t)} \right| = \mathcal{O}\left(d^2 \sqrt{\log n} \right) \right] \geqslant 1 - 2n^{-1}.$$

- *After $\mathcal{O}\left(\frac{\log(Kn)}{1-\lambda(\mathbf{P})}\right)$ rounds, the discrepancy is at most $\mathcal{O}\left(d^2\sqrt{\log n}\right)$ w. p. at least $1 - n^{-1}$.*

Proof. We prove this result in the same way as Theorem 2.14, but now we invoke Lemma 5.4 instead of Lemma 2.12. Fix any node $w \in V$, round t and define $Z_w := x_w^{(t)} - \xi_w^{(t)}$. By (5.1),

$$Z_w = x_w^{(t)} - \xi_w^{(t)} = \sum_{s=1}^{t} \sum_{[u:v]\in E} \left(\mathbf{P}_{w,u}^{t-s} - \mathbf{P}_{v,u}^{t-s} \right) \cdot e_{u,v}^{(s)}.$$

Applying Lemma 5.4, we have $\mathbf{E}\left[\, Z_w \,\right] = 0$ and for any $\delta > 0$ that

$$\mathbf{Pr}\left[\, |Z_w| \geqslant \delta \,\right] \leqslant 2\exp\left(- \frac{\delta^2}{8 \sum_{s=1}^{t} \sum_{[u:v]\in E} \left(\mathbf{P}_{w,u}^{t-s} - \mathbf{P}_{v,u}^{t-s} \right)^2} \right).$$

By the definition of the local 2-divergence, the denominator above is upper bounded by $8 \cdot \Psi_2(\mathbf{P})^2$, and we obtain for $\delta = 4\sqrt{\log n} \cdot \Psi_2(\mathbf{P})$ that

$$\mathbf{Pr}\left[\, |Z_w| \leqslant 4\sqrt{\log n} \cdot \Psi_2(\mathbf{P}) \,\right] \geqslant 1 - 2n^{-2},$$

and the first statement follows by using the union bound and the upper bound on $\Psi_2(\mathbf{P})$ from Theorem 5.3. The second statement follows directly by applying Theorem 5.1.

For the vertex-based algorithm, it was shown in [8, Proof of Thm. 1.1] that

$$\mathbf{Pr}\left[\max_{w\in V} \left| x_w^{(t)} - \xi_w^{(t)} \right| = \mathcal{O}\left(\sqrt{\log n} \cdot d \cdot \Upsilon_2(\mathbf{P}) \right) \right] \geqslant 1 - n^{-1}.$$

Using Theorem 5.3 with $\gamma = 1{+}1/d$ (and $\Delta = d$) gives $\Upsilon_2(\mathbf{P}) \leqslant \Psi_2(\mathbf{P}) = \mathcal{O}(d)$, which yields the third statement. Finally, the bound on the discrepancy follows immediately from Theorem 5.1. ∎

References

[1] M. Adler, S. Chakrabarti, M. Mitzenmacher, and L. E. Rasmussen. Parallel randomized load balancing. *Random Struct. Algorithms*, 13(2):159–188, 1998.

[2] M. Adler, E. Halperin, R. M. Karp, and V. V. Vazirani. A stochastic process on the hypercube with applications to peer-to-peer networks. In *Proc. 35th Symp. on Theory of Computing (STOC)*, pages 575–584, 2003.

[3] W. Aiello, B. Awerbuch, B. M. Maggs, and S. Rao. Approximate load balancing on dynamic and asynchronous networks. In *Proc. 25th Symp. on Theory of Computing (STOC)*, pages 632–641, 1993.

[4] M. Ajtai, J. Komlós, and E. Szemerédi. Sorting in $c \log n$ parallel steps. *Combinatorica*, 3: 1–19, 1983.

[5] J. Aspnes, M. Herlihy, and N. Shavit. Counting networks and multi-processor coordination. *J. ACM*, 41:1020–1048, 1994.

[6] C. Augonnet, S. Thibault, R. Namyst, and P.-A. Wacrenier. Starpu: a unified platform for task scheduling on heterogeneous multicore architectures. *Concurrency and Computation: Practice and Experience*, 23(2):187–198, 2011.

[7] P. Berenbrink, T. Friedetzky, and Z. Hu. A new analytical method for parallel, diffusion-type load balancing. *J. Parallel and Distributed Comput.*, 69:54–61, 2009.

[8] P. Berenbrink, C. Cooper, T. Friedetzky, T. Friedrich, and T. Sauerwald. Randomized diffusion for indivisible loads. In *Proc. 22nd Symp. on Discrete Algorithms (SODA)*, pages 429–439, 2011.

[9] J. Boillat. Load balancing and poisson equation in a graph. *Concurrency: Pract. Exper.*, 2:289–313, 1990.

[10] J. Boillat, F. Bruge, and P. Kropf. A dynamic load balancing algorithm for molecular dynamics simulation on multiprocessor systems. *J. Comp. Physics*, 96:1–14, 1991.

[11] S. P. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE/ACM Trans. Netw.*, 14(6):2508–2530, 2006.

[12] G. Cybenko. Load balancing for distributed memory multiprocessors. *J. Parallel and Distributed Comput.*, 7:279–301, 1989.

[13] D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[14] D. Dubhashi and D. Ranjan. Balls and bins: A study in negative dependence. *Random Struct. Algorithms*, 13(2):99–124, 1998.

[15] R. Elsässer and T. Sauerwald. Discrete load balancing is (almost) as easy as continuous load balancing. In *Proc. 29th Symp. Principles of Distributed Computing (PODC)*, pages 346–354, 2010.

[16] E. Even-Dar and Y. Mansour. Fast convergence of selfish rerouting. In *Proc. 16th Symp. on Discrete Algorithms (SODA)*, pages 772–781, 2005.

[17] T. Friedrich and T. Sauerwald. Near-perfect load balancing by randomized rounding. In *Proc. 41st Symp. on Theory of Computing (STOC)*, pages 121–130, 2009.

[18] T. Friedrich, M. Gairing, and T. Sauerwald. Quasirandom load balancing. In *Proc. 21st Symp. on Discrete Algorithms (SODA)*, pages 1620–1629, 2010.

[19] B. Ghosh, F. T. Leighton, B. M. Maggs, S. Muthukrishnan, C. G. Plaxton, R. Rajaraman, A. W. Richa, R. E. Tarjan, and D. Zuckerman. Tight analyses of two local load balancing algorithms. *SIAM J. Comput.*, 29(1):29–64, 1999.

[20] M. Herlihy and S. Tirthapura. Randomized smoothing networks. *J. Parallel and Distributed Comput.*, 66:626–632, 2006.

[21] D. R. Karger and M. Ruhl. Simple efficient load balancing algorithms for peer-to-peer systems. In *Proc. 16th Symp. on Parallelism in Algorithms and Architectures (SPAA)*, pages 36–43, 2004.

[22] K. Kenthapadi and R. Panigrahy. Balanced allocation on graphs. In *Proc. 17th Symp. on Discrete Algorithms (SODA)*, pages 434–443, 2006.

[23] M. Klugerman and C. G. Plaxton. Small-depth counting networks. In *Proc. 24th Symp. on Theory of Computing (STOC)*, pages 417–428, 1992.

[24] C. Lenzen and R. Wattenhofer. Tight bounds for parallel randomized load balancing: extended abstract. In *Proc. 43rd Symp. on Theory of Computing (STOC)*, pages 11–20, 2011.

[25] Y. Liu, X. Zhang, H. Li, and D. Qian. Allocating tasks in multi-core processor based parallel system. In *IFIP International Conference on Network and Parallel Computing Workshops*, pages 748–753, 2007.

[26] L. Lovász and P. Winkler. Mixing of random walks and other diffusions on a graph. *Surveys in combinatorics*, pages 119–154, 1995.

[27] R. Lyons. Asymptotic enumeration of spanning trees. *Comb., Probab. & Comput.*, 14(4): 491–522, 2005.

[28] G. S. Manku. Balanced binary trees for ID management and load balance in distributed hash tables. In *Proc. 23rd Symp. Principles of Distributed Computing (PODC)*, pages 197–205, 2004.

[29] M. Mavronicolas and T. Sauerwald. The impact of randomization in smoothing networks. *Distributed Computing*, 381-411(22), 2010.

[30] M. Mitzenmacher. *The Power of Two Choices in Randomized Load Balancing.* PhD thesis, University of California, Berkeley, 1996.

[31] S. Muthukrishnan and B. Ghosh. Dynamic load balancing by random matchings. *J. Comput. Syst. Sci.*, 53:357–370, 1996.

[32] S. Muthukrishnan, B. Ghosh, and M. H. Schultz. First- and second-order diffusive methods for rapid, coarse, distributed load balancing. *Theory Comput. Syst.*, 31(4):331–354, 1998.

[33] A. Panconesi and A. Srinivasan. Improved distributed algorithms for coloring and network decomposition problems. In *Proc. 24th Symp. on Theory of Computing (STOC)*, pages 581–592, 1992.

[34] A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM J. Comput.*, 26(2):350–368, 1997.

[35] Y. Rabani, A. Sinclair, and R. Wanka. Local divergence of Markov chains and the analysis of iterative load balancing schemes. In *Proc. 39th Symp. on Foundations of Computer Science (FOCS)*, pages 694–705, 1998.

[36] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Comb., Probab. & Comput.*, 1:351–370, 1992.

[37] A. Sinclair and M. Jerrum. Approximate counting, uninform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

[38] R. Subramanian and I. D. Scherson. An analysis of diffusive load-balancing. In *Proc. 6th Symp. on Parallelism in Algorithms and Architectures (SPAA)*, pages 220–225, 1994.

[39] S. Surana, B. Godfrey, K. Lakshminarayanan, R. Karp, and I. Stoica. Load balancing in

dynamic structured peer-to-peer systems. *Performance Evaluation*, 63(3):217–240, 2006.

[40] R. D. Williams. Performance of dynamic load balancing algorithms for unstructured mesh calculations. *Concurrency: Practice and Experience*, 3(5):457–481, 1991.

[41] C.-Z. Xu, F. C. M. Lau, B. Monien, and R. Lüling. Nearest-neighbor algorithms for load-balancing in parallel computers. *Concurrency: Pract. Exper.*, 7(7):707–736, 1995.

[42] D. Zhanga, C. Jianga, and S. Li. A fast adaptive load balancing method for parallel particle-based simulations. *Simulation Modelling Practice and Theory*, 17(6):1032–1042, 2009.

A Concentration Inequalities

The following concentration inequality is also known as "Method of Averaged Bounded Differences".

Theorem A.1 ([13, page 83])**.** *Let $Y_1, \ldots, Y_n$ be an arbitrary set of random variables and let f be a function of these random variables satisfying the property that for each $\ell \in \{1, \ldots, n\}$, there is a non-negative c_ℓ such that*

$$\left| \mathbf{E}\left[f \mid Y_\ell, Y_{\ell-1}, \ldots, Y_1 \right] - \mathbf{E}\left[f \mid Y_{\ell-1}, \ldots, Y_1 \right] \right| \leqslant c_\ell.$$

Then for any $\delta > 0$,

$$\mathbf{Pr}\left[|f - \mathbf{E}\left[f \right]| > \delta \right] \leqslant 2 \exp\left(-\frac{\delta^2}{2\sum_{\ell=1}^{n} c_\ell^2} \right).$$

The following result is borrowed from [13]. It can be shown easily using the Taylor expansion of $\mathbf{E}\left[e^{tX} \right]$.

Lemma A.2 ([13, Problem 1.14])**.** *Let $X_1, X_2, \ldots, X_n$ be independent $0/1$-random variables with $\mathbf{Pr}\left[X_i = 1 \right] = p_i$. Let $X := \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}\left[X \right] = \sum_{i=1}^{n} p_i$. If for all subsets $S \subseteq \{1, \ldots, n\}$, $\mathbf{Pr}\left[\bigwedge_{i \in S} (X_i = 1) \right] \leqslant \prod_{i \in S} \mathbf{Pr}\left[X_i = 1 \right]$, then it holds for all $\delta > 0$ that*

$$\mathbf{Pr}\left[X \geqslant (1+\delta)\mu \right] \leqslant \left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^\mu.$$

The following lemma is a standard Chernoff bound for the sum of independent, identically distributed geometric random variables.

Lemma A.3. *Consider some fixed $0 < p < 1$. Suppose that $X_1, \ldots, X_n$ are independent geometric random variables on $\mathbb{N}$ with $\mathbf{Pr}\left[X_i = k \right] = (1-p)^{k-1} p$ for every $k \in \mathbb{N}$. Let $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbf{E}\left[X \right]$. Then it holds for all $\beta > 0$ that*

$$\mathbf{Pr}\left[X \geqslant (1+\beta)\mu \right] \leqslant \exp\left(-\frac{\beta^2 \cdot n}{2\,(1+\beta)} \right).$$

We continue to define the notion of negative regression.

Definition A.4 ([14, Definition 21])**.** *A random vector $X = (X_1, \ldots, X_n) \in \{0,1\}^n$ is said to satisfy the* negative regression condition *if for any two disjoint subsets $\mathcal{I}$ and $\mathcal{J}$ of $\{1, \ldots, n\}$ and any non-decreasing function $f \colon \{0,1\}^{|\mathcal{I}|} \to \mathbb{R}$,*

$$\mathbf{E}\left[f(X_i, i \in \mathcal{I}) \;\middle|\; X_j = \sigma_j, j \in \mathcal{J} \right]$$

is non-increasing in each $\sigma_j \in \{0,1\}, j \in \mathcal{J}$.

Lemma A.5 ([14, Lemma 26])**.** *Let $X = (X_1, \ldots, X_n) \in \{0,1\}^\gamma$ be a random vector that satisfies the negative regression condition. Then for any index set $\mathcal{I} \subseteq \{1, \ldots, n\}$ and any non-decreasing functions $f_i, i \in \mathcal{I}$,*

$$\mathbf{E}\left[\prod_{i \in \mathcal{I}} f_i(X_i) \right] \leqslant \prod_{i \in \mathcal{I}} \mathbf{E}\left[f_i(X_i) \right].$$

B Useful Inequalities for Markov Chains

Lemma B.1 ([27, Lemma 3.4 and Remark 4])**.** *Let* $\mathbf{Q}$ *be the transition matrix of a reversible, ergodic Markov chain with stationary distribution* π*. Let* $a := \inf_x\{\mathbf{Q}_{x,x} \colon \mathbf{Q}_{x,x} > 0\}$ *and* $c := \inf\{\pi(x)\mathbf{Q}_{x,y} \colon x \neq y \text{ and } \mathbf{Q}_{x,y} > 0\} > 0$*. Then for any two states* x, y*, and any round* $t \geqslant 0$*,*

$$\mathbf{Q}^t_{x,y} \leqslant \pi(y) + \frac{\pi(y)}{ac\sqrt{t+1}}.$$

We now apply Lemma B.1 to the balancing circuit model where the number of matchings d is a constant. Recall that $\mathbf{M} = \prod_{i=1}^d \mathbf{M}^{(i)}$ and $\mathbf{M}^t$ is the t-th power of the matrix $\mathbf{M}$ (which is different from $\mathbf{M}^{(t)}$).

Corollary B.2. *Consider the balancing circuit model with* $d = \mathcal{O}(1)$ *matchings* $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(d)}$ *such that* $\mathbf{M} = \prod_{i=1}^d \mathbf{M}^{(i)}$ *corresponds to an ergodic Markov chain. Then for any two nodes* $u, v \in V$ *and any round* $t \geqslant 0$ *it holds that*

$$\mathbf{M}^t_{u,v} \leqslant \frac{1}{n} + \mathcal{O}\left(\frac{1}{\sqrt{t}}\right).$$

Proof. Note that $\mathbf{M}$ is a symmetric matrix which corresponds to a reversible, ergodic Markov chain with uniform stationary distribution. Further, in the notation of Lemma B.1, $a \geqslant 2^{-d}$ and $c \geqslant (1/n)2^{-d}$. Hence applying Lemma B.1 to the round matrix $\mathbf{M}$ implies

$$\mathbf{M}^t_{u,v} \leqslant \frac{1}{n} + \frac{\frac{1}{n}}{2^{-d} \cdot \frac{1}{n}2^{-d} \cdot \sqrt{t+1}} = \frac{1}{n} + \mathcal{O}\left(\frac{1}{\sqrt{t}}\right) \quad . \qquad \blacksquare$$

The next lemma is a well-known fact in Markov chain theory.

Lemma B.3 ([32, Lemma 1])**.** *Consider the balancing circuit model for* d *matchings* $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(d)}$*. Let* $\mathbf{M} = \prod_{i=1}^d \mathbf{M}^{(i)}$*. Then, for any* $u \in V$ *it holds that*

$$\sum_{v \in V} \left(\mathbf{M}^t_{u,v} - \frac{1}{n}\right)^2 \leqslant \lambda^{2t}.$$

The following lemma is the corresponding result of Corollary B.2 for the random matching model.

Lemma B.4. *Let* G *be any* d*-regular graph and consider the random matching model. Fix any node* $u \in V$*. Then, for any* σ *with* $\frac{2}{n} \leqslant \sigma^{-1} \leqslant 1$*, there is a constant* c *independent of* σ *and* n *such that*

$$\mathbf{Pr}\left[\left\|\mathbf{M}^{[1,c\sigma^6]}_{u,\cdot}\right\|_2^2 \geqslant \sigma^{-1}\right] \leqslant \mathrm{e}^{-\sigma^4}.$$

Proof. Consider a round s for which $\left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_2^2 \geqslant \sigma^{-1}$. Since $\|z\|_2^2 \leqslant \|z\|_1 \cdot \|z\|_\infty$ holds for any vector z, we have

$$\left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_\infty \geqslant \frac{\left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_2^2}{\left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_1} \geqslant \sigma^{-1}.$$

Let v be any node with $\mathbf{M}^{[1,s]}_{u,v} \geqslant \sigma^{-1}$. We continue with a case distinction on the degree of G.

Case 1: $d \geqslant 4\sigma^2$. Since $\sum_{w\in N(v)} \mathbf{M}^{[1,s]}_{u,w} \leqslant 1$, at least $d/2$ of the neighbors $w \in N(v)$ satisfy $\mathbf{M}^{[1,s]}_{u,w} \leqslant \frac{1}{2}\sigma^{-2}$. Hence an edge $\{v,w\}$ with $\mathbf{M}^{[1,s]}_{u,w} \leqslant \frac{1}{2}\sigma^{-2}$ is included in the random matching in round $s+1$ with constant probability. Assuming that this event happens, then

$$\begin{aligned}
\left\|\mathbf{M}^{[1,s+1]}_{u,\cdot}\right\|_2^2 &\leqslant \sum_{k\in V,k\not\in\{v,w\}} \left(\mathbf{M}^{[1,s]}_{u,k}\right)^2 + 2\cdot\left(\frac{\mathbf{M}^{[1,s]}_{u,v}+\mathbf{M}^{[1,s]}_{u,w}}{2}\right)^2 \\
&= \left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_2^2 - \left(\mathbf{M}^{[1,s]}_{u,v}\right)^2 - \left(\mathbf{M}^{[1,s]}_{u,w}\right)^2 + \frac{1}{2}\left(\mathbf{M}^{[1,s]}_{u,v}+\mathbf{M}^{[1,s]}_{u,w}\right)^2 \\
&= \left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_2^2 - \frac{1}{2}\left(\mathbf{M}^{[1,s]}_{u,v}-\mathbf{M}^{[1,s]}_{u,w}\right)^2 \qquad \text{(B.1)} \\
&\leqslant \left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_2^2 - \frac{1}{2}\left(\sigma^{-1}-\frac{1}{2}\sigma^{-2}\right)^2 \\
&\leqslant \left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_2^2 - \frac{1}{8}\sigma^{-2}.
\end{aligned}$$

Case 2. $d \leqslant 4\sigma^2$. Since $\frac{1}{2}\sigma^{-1} \geqslant \frac{1}{n}$ by the assumption on σ, there is at least one node w with $\mathbf{M}^{[1,s]}_{u,w} \leqslant \frac{1}{2}\sigma^{-1}$. Consider now a shortest path $P = (u_1 = u,\ldots,u_\ell = w)$ from u to such a node w with the property that w is the first node on P with $\mathbf{M}^{[1,s]}_{u,w} \leqslant \frac{1}{2}\sigma^{-1}$. By construction of P, the length ℓ satisfies $\ell - 1 \leqslant 1/(\frac{1}{2}\sigma^{-1}) = 2\sigma$. Hence, there must be at least one edge $\{f,g\} \in E$ along the path P so that

$$\mathbf{M}^{[1,s]}_{u,f} - \mathbf{M}^{[1,s]}_{u,g} \geqslant \frac{\mathbf{M}^{[1,s]}_{u,v} - \mathbf{M}^{[1,s]}_{u,w}}{\ell-1} \geqslant \frac{\frac{1}{2}\sigma^{-1}}{2\sigma} = \frac{1}{4}\sigma^{-2}.$$

Note that the edge $\{f,g\}$ is included in the matching in round $s+1$ with probability $\Omega(1/d) = \Omega(\sigma^{-2})$. If the edge $\{f,g\}$ is part of the matching, then we conclude from (B.1) that

$$\left\|\mathbf{M}^{[1,s+1]}_{u,\cdot}\right\|_2^2 \leqslant \left\|\mathbf{M}^{[1,s]}_{u,\cdot}\right\|_2^2 - \frac{1}{32}\sigma^{-4}.$$

Summarizing both cases, we can upper bound the minimum round τ before $\left\|\mathbf{M}^{[1,\tau]}_{u,\cdot}\right\|_2^2 \leqslant \sigma^{-1}$ occurs by the sum of $32\sigma^4$ independent random geometric variables with success probability $\Omega(d^{-1}) = \Omega(\sigma^{-2})$ each. Using Lemma A.3, we obtain that the sum of these geometric variables is larger than $c\cdot\sigma^6$ with probability at most $\mathrm{e}^{-\sigma^4}$, if c is a sufficiently large constant. ■

Lemma B.5. *Fix any sequence of matchings $\mathcal{M} = \langle \mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \ldots\rangle$ and consider the continuous process. Assume that the time-interval $[0,t]$ is (K,ε)–smoothing. Then for any non-negative vector y with $\|y\|_1 = 1$ it holds that $\left|\sum_w y_w \xi^{(t)}_w - \overline{\xi}\right| \leqslant \varepsilon$. Moreover, for any $t \geqslant \tau_{\mathrm{cont}}(1,\varepsilon)$, $\left|\mathbf{M}^{[1,t]}_{u,v} - \frac{1}{n}\right| \leqslant \varepsilon$.*

Proof. Let $\xi^{(0)}$ be any load vector with initial discrepancy at most K. Since $[0,t]$ is (K,ε)-smoothing, $\mathrm{disc}(\xi^{(t)}) \leqslant \varepsilon$. That is, for all pairs of nodes $u,v \in V$ it holds that

$$\left|\xi^{(t)}_u - \xi^{(t)}_v\right| \leqslant \varepsilon.$$

Consequently, for all nodes $w \in V$,

$$\left|\xi^{(t)}_w - \overline{\xi}\right| \leqslant \varepsilon.$$

Using the above inequality along with the triangle inequality, we get

$$\begin{aligned}\left|\sum_{w\in V} y_w\xi_w^{(t)} - \overline{\xi}\right| &= \left|\sum_{w\in V}\left(y_w\xi_w^{(t)} - y_w\overline{\xi}\right)\right| \\ &\leqslant \sum_{w\in V}\left|y_w\xi_w^{(t)} - y_w\overline{\xi}\right| \\ &= \sum_{w\in V} y_w\cdot\left|\xi_w^{(t)} - \overline{\xi}\right| \leqslant \sum_{w\in V} y_w\cdot\varepsilon = \varepsilon.\end{aligned}$$

For the second statement, let y be the unit-vector which is one at component v and $\xi^{(0)}$ be the unit-vector which is one at component u. Then $\overline{\xi} = \frac{1}{n}$, $\xi_v^{(t)} = \sum_{w\in V}\xi_w^{(0)}\mathbf{M}_{w,v}^{[1,t]} = \mathbf{M}_{u,v}^{[1,t]}$. Hence $\sum_{w\in V} y_w\xi_w^{(t)} = \xi_v^{(t)} = \mathbf{M}_{u,v}^{[1,t]}$ and $\left|\mathbf{M}_{u,v}^{[1,t]} - \frac{1}{n}\right| \leqslant \varepsilon$. ■